

Query Match 3.6%; Score 7; DB 2; Length 729;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 IKKEQDK 177

|||||

Db 359 IKKEQDK 365

RESULT 37

I59282

diacylglycerol kinase (EC 2.7.1.107) gamma - rat

C:Species: Rattus sp. (rat)

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C:Accession: I59282

R:Goto, K.; Funayama, M.; Kondo, H.

A:Title: Cloning and expression of a cytoskeleton-associated diacylglycerol kinase that

A:Reference number: I59282; MUID:95108095; PMID:7809169

A:Accession: I59282

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-788 <RES>

A:Cross-references: GB:D38448; NID:9643598; PIDN:BAA07480.1; PID:9784935

A:Superfamily: human diacylglycerol kinase; calmodulin repeat homology; protein kinase C

C:Keywords: ATP; calcium binding; duplication; EF hand; phosphotransferase; zinc

F:172-204/Domain: calmodulin repeat homology <EF1>

F:217-249/Domain: calmodulin repeat homology <EF2>

F:269-318/Domain: protein kinase C zinc-binding repeat homology <KZ1>

F:334-380/Domain: protein kinase C zinc-binding repeat homology <KZ2>

Query Match

Best Local Similarity 100.0%; Pred. No. 48;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 RRKRSRK 134

|||||

Db 781 RRKRSRK 787

RESULT 38

A53691

diacylglycerol kinase (EC 2.7.1.107) gamma - human

C:Species: Homo sapiens (man)

C:Date: 07-Oct-1994 #sequence_revision 07-Oct-1994 #text_change 20-Jun-2000

C:Accession: A53691

R:Kai, M.; Sakane, F.; Imai, S.; Wada, I.; Kanoh, H.

J. Biol. Chem. 269, 18492-18498, 1994

A:Title: Molecular cloning of a diacylglycerol kinase isozyme predominantly expressed in

A:Reference number: A53691; MUID:94308084; PMID:8034597

A:Accession: A53691

A:Molecule type: mRNA

A:Residues: 1-791 <KAI>

A:Cross-references: GB:D26135; NID:9516757; PIDN:BAA05132.1; PID:9516758

C:Genetics:

A:Gene: GDB:DAGK3

A:Cross-references: GDB:377343

A:Map position: 12q13.3-12q13.3

C:Superfamily: human diacylglycerol kinase; calmodulin repeat homology; protein kinase C

C:Keywords: ATP; calcium binding; duplication; EF hand; phosphotransferase; zinc

F:175-207/Domain: calmodulin repeat homology <EF1>

F:220-252/Domain: calmodulin repeat homology <EF2>

F:272-321/Domain: protein kinase C zinc-binding repeat homology <KZ1>

F:337-383/Domain: protein kinase C zinc-binding repeat homology <KZ2>

Query Match

Best Local Similarity 100.0%; Pred. No. 48;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 RRKRSRK 134

|||||

Db 784 RRKRSRK 790

RESULT 39

JN0878

100K protein - fowl adenovirus 10

C:Species: Aviadenovirus gallio (fowl adenovirus 10)

C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 08-Oct-1999

C:Accession: JN0878

R:Sheppard, M.

Gene 132, 307-308, 1993

A:Title: Identification of a fowl adenovirus gene with sequence homology to the 100K

A:Reference number: JN0878; MUID:94040780; PMID:8224879

A:Accession: JN0878

A:Molecule type: DNA

A:Residues: 1-798 <SHE>

A:Cross-references: GB:L07890; NID:9348034; PIDN:AAA72328.1; PID:9348035

A:Experimental source: serotype 10

C:Genetics:

A:Gene: 100K

A:Map position: 57.3 to 62.6 map units

C:Superfamily: adenovirus late 100K protein

Query Match

Best Local Similarity 100.0%; Pred. No. 48;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 ENPNRPD 117

|||||

Db 592 ENPNRPD 598

RESULT 40

C84615

hypothetical protein At2g22660 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C:Accession: C84615

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: C84615

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-819 <STO>

A:Cross-references: GB:AE002093; NID:94314363; PIDN:RAD15574.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g22660

A:Map position: 2

Query Match

Best Local Similarity 100.0%; Pred. No. 49;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 182 AKOHLFF 188

|||||

Db 27 AKOHLFF 33

Search completed: February 26, 2003, 10:01:52

Job time : 71 secs

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 VTPEMLN 11
|||||||
Db 68 VTPEMLN 74

RESULT 32

S64250

Probable membrane protein YGL228w - yeast (*Saccharomyces cerevisiae*)

N:Alternate names: hypothetical protein G0954

C:Species: *Saccharomyces cerevisiae*

C>Date: 17-May-1996 #sequence_revision 17-May-1996 #text_change 19-Apr-2002

C:Accession: S64250

R:Rattmann, B.; Kramer, B.; Kramer, W.

Submitted to the Protein Sequence Database, May 1996

A:Reference number: S64248

A:Accession: S64250

A:Molecule type: DNA

A:Residues: 1-577 <FAR>

A:Cross-references: EMBL:Z72750; NID:g1322882; PID:e243889; PID:g1322883; GSPDB:GN000007;

A:Experimental source: strain S288C

C:Genetics:

A:Gene: SGD:SHE10; MIPS:YGL228w

A:Cross-references: SGD:S0003197

A:Map position: 7L

C:Keywords: transmembrane protein

F:2-18/Domain: transmembrane status predicted <TM>

Query Match

Best Local Similarity 3.6%; Score 7; DB 2; Length 577;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 KYDFIVG 65
|||||||
Db 175 KYDFIVG 181

RESULT 33

T16774

hypothetical protein SSSD1.1 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

C:Accession: T16774

R:Minx, P.

Submitted to the EMBL Data Library, November 1995

A:Description: The sequence of *C. elegans* cosmid SSSD1.

A:Reference number: Z18575

A:Accession: T16774

A:Molecule type: DNA

A:Residues: 1-599 <MIN>

A:Cross-references: EMBL:U41530; NID:g1118030; PID:g1118031; PIDN:AA03273.1; CESP:SSSD1

C:Genetics:

A:Gene: CESP:SSSD1.1

A:Introns: 28/1; 67/3; 107/3; 125/3; 142/2; 180/3; 230/2; 262/1; 302/3; 343/3; 370/1; 38

Query Match

Best Local Similarity 3.6%; Score 7; DB 2; Length 599;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 KRRKRS 133
|||||||
Db 575 KRRKRS 581

RESULT 34

B86789

hypothetical protein pabB [imported] - *Lactococcus lactis* subsp. *lactis* (strain ILJ403)

C:Species: *Lactococcus lactis* subsp. *lactis*

C>Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001

C:Accession: B86789

R:Boilotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarre, K.; Weissenbach, J.; Ehrlich

Genome Res. 11, 731-753, 2001

A:Title: The complete genome sequence of the lactic acid bacterium *Lactococcus lactis*

A:Reference number: A86625; MUID:21235186; PMID:11337471

A:Accession: B86789

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-628 <STO>

A:Cross-references: GB:AE005176; PID:g12724293; PIDN:AAK05412.1; GSPDB:GN00146

A:Experimental source: strain ILJ403

C:Genetics:

A:Gene: pabB

Query Match

Best Local Similarity 3.6%; Score 7; DB 2; Length 628;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 FSEILK 59
|||||||
Db 413 FSEILK 419

RESULT 35

F84996

phosphotransferase system enzyme II (EC 2.7.1.69) [imported] - *Buchnera* sp. (strain A

N:Alternate names: pts system mannitol-specific IIABC component

C:Species: *Buchnera* sp.

C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001

C:Accession: F84996

R:Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.

Nature 407, 81-86, 2000

A:Title: Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp

A:Reference number: A84930; MUID:20445173; PMID:10993077

A:Accession: F84996

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-632 <STO>

A:Cross-references: GB:AP000398; GSPDB:GN00144

A:Experimental source: strain APS

C:Genetics:

A:Gene: mtiA; B0572

C:Superfamily: phosphotransferase system mannitol-specific enzyme II; phosphotransfer

actor III homology

C:Keywords: phosphotransferase

Query Match

Best Local Similarity 3.6%; Score 7; DB 2; Length 632;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 ILLKYDF 62
|||||||
Db 332 ILLKYDF 338

RESULT 36

A97747

oligopectidase B (EC 3.4.21.83) [imported] - *Rickettsia conorii* (strain Malish 7)

C:Species: *Rickettsia conorii*

C>Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 17-May-2002

C:Accession: A97747

R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.;

Science 293, 2093-2098, 2001

A:Title: Mechanisms of Evolution in *Rickettsia conorii* and *Rickettsia prowazekii*.

A:Reference number: A97700; MUID:21442074; PMID:11557893

A:Accession: A97747

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-729 <KUR>

A:Cross-references: GB:AE006914; PIDN:AAI02915.1; PID:g15619442; GSPDB:GN00173

C:Genetics:

A:Gene: ptrB

C:Superfamily: prolyl oligopeptidase

C:Keywords: hydrolase; serine proteinase

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 EKSAPDV 46

|||||||

Db 347 EKSAPDV 353

RESULT 28

T20382

hypothetical protein D2089.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999

C:Accession: T20382; T23815

R:Swinsburne, J.

submitted to the EMBL Data Library, September 1994

A:Reference number: Z19264

A:Status: preliminary; translated from GB/EMBL/DBBJ

A:Accession: T20382

A:Molecule type: DNA

A:Residues: 1-480 <WIL>

A:Cross-references: EMBL:Z36948; PIDN:CAA85414.1; GSPDB:GN00020; CESP:D2089.1

A:Experimental source: clone D2089

R:Gardner, A.

submitted to the EMBL Data Library, June 1995

A:Reference number: Z19803

A:Accession: T23815

A:Status: preliminary; translated from GB/EMBL/DBBJ

A:Molecule type: DNA

A:Residues: 1-480 <W12>

A:Cross-references: EMBL:Z49911; PIDN:CAA90135.1; GSPDB:GN00020; CESP:D2089.1

A:Experimental source: clone M28

C:Genetics:

A:Gene: CESP:D2089.1

A:Map position: 2

A:Introns: 40/3; 91/1; 132/1; 158/3; 217/3; 262/2

Query Match

Best Local Similarity 3.6%; Score 7; DB 2; Length 480;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 IKFDDER 124

|||||||

Db 59 IKFDDER 65

RESULT 29

S67625

probable membrane protein YDL089w - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein D2416

C:Species: Saccharomyces cerevisiae

C>Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 19-Apr-2002

C:Accession: S67625; S67631; S67428

R:Wambutt, R.; Wedler, H.; Wedler, E.; Scharfe, M.

submitted to the Protein Sequence Database, July 1996

A:Reference number: S67608

A:Accession: S67625

A:Molecule type: DNA

A:Residues: 1-484 <WAM>

A:Cross-references: EMBL:Z74137; NID:g1431117; PID:g1431118; MIPS:YDL089w

A:Experimental source: strain S288C

R:Bailestia, J.P.G.; Remacha, M.; Soler-Mira, A.; Jimenez, A.; Garcia-Cantalejo, J.M.; Bo

submitted to the Protein Sequence Database, July 1996

A:Reference number: S67629

A:Accession: S67631

A:Molecule type: DNA

A:Residues: 1-484 <BAL>

A:Cross-references: EMBL:Z74137; NID:g1431117; PID:g1431118; MIPS:YDL089w

A:Experimental source: strain S288C

R:Boskovic, J.; Saiz, J.E.; Soler-Mira, A.; Jimenez, A.; Garcia-Cantalejo, J.; Revuelta, J.L.; Jimine

submitted to the EMBL Data Library, February 1996

A:Reference number: S67406

A:Accession: S67428

A:Molecule type: DNA

A:Residues: 1-484 <BOS>

A:Cross-references: EMBL:X95644; NID:g1199535; PID:g1199558

C:Genetics:

A:Cross-references: SGD:S0002247

A:Map position: 4L

A>Note: YDL089w

C:Superfamily: Saccharomyces cerevisiae probable membrane protein YDL089w

C:Keywords: transmembrane protein

F:68-84/Domain: transmembrane #status predicted <TM1>

F:147-163/Domain: transmembrane #status predicted <TM2>

F:241-257/Domain: transmembrane #status predicted <TM3>

Query Match

Best Local Similarity 3.6%; Score 7; DB 2; Length 484;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 QLVINEK 41

|||||||

Db 42 QLVINEK 48

RESULT 30

T29814

hypothetical protein C46A5.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T29814

R:Johnson, D.; Stellyes, L.

submitted to the EMBL Data Library, June 1996

A:Description: The sequence of C. elegans cosmid C46A5.

A:Reference number: Z20690

A:Accession: T29814

A:Status: preliminary; translated from GB/EMBL/DBBJ

A:Molecule type: DNA

A:Residues: 1-497 <JOH>

A:Cross-references: EMBL:U61948; PIDN:AA03142.1; GSPDB:GN00022; CESP:C46A5.1

A:Experimental source: strain Bristol N2; clone C46A5

C:Genetics:

A:Gene: CESP:C46A5.1

A:Map position: 4

A:Introns: 85/3; 108/1; 189/1; 221/3; 329/2; 367/2; 444/3; 473/3

Query Match

Best Local Similarity 3.6%; Score 7; DB 2; Length 497;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 RKSRSKS 135

|||||||

Db 35 RKSRSKS 41

RESULT 31

A87627

PDZ domain family protein [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001

C:Accession: A87627

R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,

n, J.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.

A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: A87627

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-561 <STO>

A:Cross-references: GB:AE005673; NID:g13424697; PIDN:AAK25013.1; GSPDB:GN00148

C:Genetics:

A:Gene: CC3051

Query Match

Best Local Similarity 3.6%; Score 7; DB 2; Length 561;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

C>Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 20-Sep-1999
C:Accession: S03762
R:Britton, P.; Carmenes, R.S.; Page, K.W.; Garwes, D.J.; Parra, F.
Mol. Microbiol. 2, 89-99, 1988
A:Title: Sequence of the nucleoprotein gene from a virulent British field isolate of tra
A:Reference number: S03762; MUID:86216185; PMID:2835592
C:Accession: S03762
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-382 <BRI>
A:Cross-references: GB:Y00542; NID:g59012; PIDN:CAA68607.1; PID:g59013
C:Superfamily: coronavirus nucleocapsid protein
C:Keywords: nucleocapsid

Query Match 3.6%; Score 7; DB 2; Length 382;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 129 RKSRKS 135
|||||
Db 342 RKSRKS 348

RESULT 24

E36607
nucleocapsid protein - porcine respiratory virus
N:Alternate names: N protein
C:Species: porcine respiratory virus
C>Date: 30-Aug-1991 #sequence_revision 30-Aug-1991 #text_change 20-Sep-1999
C:Accession: E36607; S10447
R:Rasschaert, D.; Duarte, M.; Laude, H.
J. Gen. Virol. 71, 2599-2607, 1990
A:Title: Porcine respiratory coronavirus differs from transmissible gastroenteritis virus
A:Reference number: A36607; MUID:91073120; PMID:2174956
C:Accession: E36607
A:Status: preliminary
A:Molecule type: genomic RNA
A:Residues: 1-382 <RAS>
A:Cross-references: GB:224675; NID:g395057; PIDN:CAA80841.1; PID:g395062
A:Experimental source: strain RM4
R:Page, K.W.; Britton, P.; Boursnell, M.E.G.
submitted to the EMBL Data Library, March 1990
A:Reference number: S10447
A:Accession: S10447
A:Molecule type: mRNA
A:Residues: 1-23 <PAC>
A:Cross-references: EMBL:X52668; NID:g58981; PIDN:CAA36895.1; PID:g58982
C:Superfamily: coronavirus nucleocapsid protein
C:Keywords: nucleocapsid

Query Match 3.6%; Score 7; DB 2; Length 382;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 129 RKSRKS 135
|||||
Db 342 RKSRKS 348

RESULT 25

S47428
nucleocapsid protein - porcine transmissible gastroenteritis virus
N:Alternate names: N protein
C:Species: porcine transmissible gastroenteritis virus
C>Date: 23-Nov-1994 #sequence_revision 19-Apr-1996 #text_change 20-Sep-1999
C:Accession: S65856; S47428
R:Chen, C.M.; Cavanagh, D.; Britton, P.
Virus Res. 38, 83-89, 1995
A:Title: Cloning and sequencing of a 8.4-kb region from the 3'-end of a Taiwanese virul
A:Reference number: S65850; MUID:96060227; PMID:8546012
A:Accession: S65856
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: genomic RNA

A:Residues: 1-382 <CH2>
A:Cross-references: EMBL:Z35758; NID:g529246; PIDN:CAA84811.1; PID:g529253
A:Experimental source: Taiwanese field isolate
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994
C:Genetics:
A:Gene: N
C:Superfamily: coronavirus nucleocapsid protein
C:Keywords: nucleocapsid

Query Match 3.6%; Score 7; DB 2; Length 382;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 129 RKSRKS 135
|||||
Db 342 RKSRKS 348

RESULT 26

F75596
hypothetical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: F75596
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: F75596
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-405 <WHI>
A:Cross-references: GB:AE001862; GB:AE001825; NID:g6460468; PIDN:AAF12358.1; PID:g646
A:Experimental source: strain R1
C:Genetics:
A:Gene: DRA0036
A:Map position: 2

Query Match 3.6%; Score 7; DB 2; Length 405;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 54 SEILKY 60
|||||
Db 234 SEILKY 240

RESULT 27

B70185
probable hydroxymethylglutaryl-CoA synthase (EC 4.1.3.5) - Lyme disease spirochete
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C>Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 05-May-2000
C:Accession: B70185
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vu
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98065943; PMID:9403685

A:Accession: B70185

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-407 <KLE>

A:Cross-references: GB:AE001169; GB:AE000783; NID:g2688611; PIDN:AAC67034.1; PID:g268

A:Experimental source: strain B31

C:Superfamily: hydroxymethylglutaryl-CoA synthase

C:Keywords: carbon-carbon lyase; coenzyme A; oxo-acid-lyase

Query Match 3.6%; Score 7; DB 2; Length 407;
Best Local Similarity 100.0%; Pred. No. 27;

F;111-173/Domain: ribonucleoprotein repeat homology <RRM2>

Query Match 3.6%; Score 7; DB 2; Length 344;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 RKSRKS 135
|||||

Db 241 RKSRKS 247

RESULT 19

AC3224 hypothetical protein Atu523 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)

C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002

C:Accession: AC3224
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, H.

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell

; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,

ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; PMID:11743193
A:Accession: AC3224

A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-364 <KUR>
A:Cross-references: GB:AE008687; PIDN:AAL46209.1; PID:g17743983; GSPDB:GN00188

A:Experimental source: strain C58 (Dupont)
C:Genetics:

A:Gene: Atu523
A:Genome: plasmid

Query Match 3.6%; Score 7; DB 2; Length 364;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 LRLRGFY 78
|||||

Db 117 LRLRGFY 123

RESULT 20

I39939 phenylalanine dehydrogenase (EC 1.4.1.-) - Bacillus sphaericus

C:Species: Bacillus sphaericus
C:Date: 19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change 11-Jun-1999

C:Accession: I39939
R:Okazaki, N.; Hibino, Y.; Asano, Y.; Ohmori, M.; Numao, N.; Kondo, K.

Gene 63, 337-341, 1988
A:Title: Cloning and nucleotide sequencing of phenylalanine dehydrogenase gene of Bacill

A:Reference number: I39939; MUID:88255879; PMID:2838396
A:Accession: I39939

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA

A:Residues: 1-381 <RES>
A:Cross-references: GB:M26661; NID:g341510; PIDN:AAA22646.1; PID:g529017

C:Genetics:
A:Gene: pdh

C:Superfamily: leucine dehydrogenase
C:Keywords: oxidoreductase

Query Match 3.6%; Score 7; DB 2; Length 381;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 IVKSDDI 32
|||||

Db 236 IVKSDDI 242

RESULT 21

VH1PC

nucleocapsid protein - porcine transmissible gastroenteritis virus

C:Species: porcine transmissible gastroenteritis virus

C:Date: 04-Dec-1986 #sequence_revision 04-Dec-1986 #text_change 29-Oct-1999

C:Accession: A04025; S03936
R:Kapke, P.A.; Brian, D.A.

Virolgy 151, 41-49, 1986
A:Title: Sequence analysis of the porcine transmissible gastroenteritis coronavirus n

A:Reference number: A94340; MUID:86181608; PMID:3008432
A:Accession: A04025

A:Molecule type: genomic RNA
A:Residues: 1-382 <KAP>

A:Cross-references: GB:M14878; NID:g335209; PIDN:AAA47915.1; PID:g335211

A:Experimental source: strain Purdue
R:Rasschaert, D.; Gelfi, J.; Laude, H.

Biochimie 69, 591-600, 1987
A:Title: Enteric coronavirus TGEV: partial sequence of the genomic RNA, its organizat

A:Reference number: S01738; MUID:88078100; PMID:2825819
A:Accession: S03936

A:Molecule type: mRNA
A:Residues: 1-236, 'w', 238-375, 'N', 377-382 <RAS>

A:Cross-references: EMBL:X06371; NID:g58995; PIDN:CAA29674.1; PID:g59001

A:Experimental source: strain Purdue-115
C:Superfamily: coronavirus nucleocapsid protein

C:Keywords: glycoprotein; nucleocapsid
F:134,154,172,267,364/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 3.6%; Score 7; DB 1; Length 382;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 RKSRKS 135
|||||

Db 342 RKSRKS 348

RESULT 22

S24282 nucleocapsid protein - porcine respiratory virus (strain 86/137004)

N:Alternate names: N protein

C:Species: porcine respiratory virus

A:Variety: strain 86/137004
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Jul-1999

C:Accession: S24282; S21311
R:Britton, P.; Mawditt, K.L.; Page, K.W.

Virus Res. 21, 181-198, 1991
A:Title: The cloning and sequencing of the virion protein genes from a British isolat

A:Reference number: S24279; MUID:92116634; PMID:1662846
A:Accession: S24282

A:Molecule type: genomic RNA
A:Residues: 1-382 <BR1>

A:Cross-references: EMBL:X60056; NID:g61342; PIDN:CAA42657.1; PID:g61346; EMBL:X55980

A:Experimental source: strain 86/137004
C:Genetics:

A:Gene: N
C:Superfamily: coronavirus nucleocapsid protein

C:Keywords: glycoprotein; nucleocapsid
F:134,154,172,267,364/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 3.6%; Score 7; DB 1; Length 382;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 RKSRKS 135
|||||

Db 342 RKSRKS 348

RESULT 23

S03762 nucleocapsid protein - porcine transmissible gastroenteritis virus (strain FS772/70)

C:Species: porcine transmissible gastroenteritis virus

C:Genetics:
A:Gene: aq_1343

Query Match 3.6%; Score 7; DB 2; Length 160;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 92 RLEDYIK 98
|||||||
Db 151 RLEDYIK 157

RESULT 14

E71407
hypothetical protein - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
A:Variety: columbia
C:Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 05-Dec-1998
C:Accession: E71407
R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirks, P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenecker, T.; Pohl, T.M.; Terry, N.; Giel, vanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B. Nature 391, 485-488, 1998
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech, erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Ans, C.; Chalwatzis, N.
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thaliana
A:Reference number: A71400; MUID:98121113; PMID:9461215
A:Accession: E71407
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-194 <BEV>
A:Cross-references: GB:257336; NID:g2244788; PID:e326896; PID:g2244808
C:Genetics:
A:Map position: 4COP9-4G3845

Query Match 3.6%; Score 7; DB 2; Length 194;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 123 ERPHRR 129
|||||||
Db 160 ERPHRR 166

RESULT 15

T05238
hypothetical protein F18A5.80 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Jul-1999
C:Accession: T05238
R:Bevan, M.; Weber, N.; Grueninger, D.; Schmidheini, T.; Bancroft, I.; Mewes, H.W.; Maye submitted to the Protein Sequence Database, February 1999
A:Reference number: Z15405
A:Accession: T05238
A:Molecule type: DNA
A:Residues: 1-194 <BEV>
A:Cross-references: EMBL:AL035528
A:Experimental source: cultivar Columbia; BAC clone F18A5
C:Genetics:
A:Map position: 4
A:Introns: 13/2; 108/1; 154/1
A:Note: F18A5.80

Query Match 3.6%; Score 7; DB 2; Length 194;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 156 TSKRRKD 162.
|||||||
Db 98 TSKRRKD 104

RESULT 16

T47871
hypothetical protein T4C21.10 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T47871
R:Choisne, N.; Robert, C.; Brottier, P.; Wincker, P.; Cattolico, L.; Artiguenave, F.; W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
submitted to the Protein Sequence Database, March 2000

A:Reference number: 224479
A:Accession: T47871
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-250 <CHO>
A:Cross-references: EMBL:AL162295
A:Experimental source: cultivar Columbia; BAC clone T4C21
C:Genetics:
A:Map position: 3
A:Introns: 34/1; 57/3; 85/1; 123/3; 171/3; 187/3; 213/3
A:Note: T4C21.10

Query Match 3.6%; Score 7; DB 2; Length 250;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 KEVTPEM 9
|||||||
Db 117 KEVTPEM 123

RESULT 17

JE0175
freezled protein-1b - human
C:Species: Homo sapiens (man)
C:Date: 03-Jul-1998 #sequence_revision 10-Jul-1998 #text_change 07-May-1999
C:Accession: JE0175
R:Hu, E.; Zhu, Y.; Fredrickson, T.; Barnes, M.; Kelsell, D.; Beeley, L.; Brooks, D. Biochem. Biophys. Res. Commun. 247, 287-293, 1998
A:Title: Tissue restricted expression of two human frzbs in preadipocytes and pancrea
A:Reference number: JE0174; MUID:98308108; PMID:9642118
A:Accession: JE0175
A:Molecule type: mRNA
A:Residues: 1-317 <HUA>
C:Genetics:
A:Gene: hFRP-1b
A:Map position: 5ql4.3-ql2.1

Query Match 3.6%; Score 7; DB 2; Length 317;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 159 KRKOTKR 165
|||||||
Db 236 KRKOTKR 242

RESULT 18

S59043
spilling factor SRp55 - human
C:Species: Homo sapiens (man)
C:Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 28-May-1999
C:Accession: S59043
R:Scraton, G.R.; Caceres, J.F.; Mayeda, A.; Bell, M.V.; Plebanski, M.; Jackson, D.G. EMBO J. 14, 4336-4349, 1995
A:Title: Identification and characterization of three members of the human SR family
A:Reference number: S59042; MUID:96016206; PMID:7556075
A:Accession: S59043
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-344 <SCR>
A:Cross-references: GB:U30883; NID:g1049087; PIDN:AAA93073.1; PID:g1049088
C:Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleopro
F:3-62/Domain: ribonucleoprotein repeat homology <RRM1>

A:Accession: S64928
A:Molecule type: DNA
A:Residues: 1-502 <BEN>
A:Cross-references: EMBL:273266; NID:g1360470; PIDN:CAA97656.1; PID:g1360471; GSPDB:GN0004
A:Experimental source: strain S288C
C:Genetics:
A:Gene: SGD:GIS3; MIPS:YLR094c
A:Cross-references: SGD:S0004084
A:Map position: 12R
C:Superfamily: Saccharomyces cerevisiae hypothetical protein YLR094c

Query Match 4.2%; Score 8; DB 2; Length 502;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 RRKRSRKS 135
|||||
Db 281 RRKRSRKS 288

RESULT 10
A57459
ribosomal protein S6 kinase 2 (EC 2.7.1.-) 2 - human
N:Alternate names: growth factor-stimulated protein kinase (GFSPK); p90 S6 kinase; ribosomal protein S6
C:Species: Homo sapiens (man)
C:Date: 16-Feb-1996 #sequence_revision 16-Feb-1996 #text_change 18-Jun-1999
C:Accession: A57459; I64834
R:Zhao, Y.; Bjorbaek, C.; Weremowicz, S.; Morton, C.C.; Moller, D.E.
Mol. Cell. Biol. 15, 4353-4363, 1995
A:Title: RSK3 encodes a novel pp90rsk isoform with a unique N-terminal sequence: growth factor-stimulated protein kinase (GFSPK)
A:Reference number: A57459; MUID:95349602; PMID:7623830
A:Accession: A57459
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-733 <ZHA>
A:Cross-references: EMBL:X85106; NID:g1033032; PIDN:CAA59427.1; PID:g1033033
R:Moller, D.E.; Xia, C.H.; Tang, W.; Zhu, A.X.; Jakubowski, M.
Am J. Physiol. 266, 351-359, 1994
A:Title: Human rsk isoforms: cloning and characterization of tissue-specific expression.
A:Reference number: I51901
A:Accession: I64834
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 'A', 1-255, 'A', 257-268, 'S', 270-446, 'G', 448-539 <MOL>
A:Cross-references: GB:I07598; NID:g401771; PIDN:AAC82496.1; PID:g401772
C:Comment: Although ribosomal protein S6 (see PIR:R3HU6) is phosphorylated by this enzyme
C:Genetics:
A:Gene: GDB:RPS6KA2; PP90RSK3; RSK; HU-2; RSK3
A:Cross-references: GDB:365645; OMIM:601685
A:Map position: 6q27-6q27
C:Superfamily: ribosomal protein S6 kinase II; protein kinase homology
C:Keywords: ATP; autophosphorylation; monomer; phosphoprotein; phosphotransferase; serine/threonine kinase
F:57-318/Domain: protein kinase homology <KIN1>
F:65-73/Region: protein kinase ATP-binding motif
F:413-672/Domain: protein kinase homology <KIN2>
F:421-429/Region: protein kinase ATP-binding motif

Query Match 4.2%; Score 8; DB 2; Length 733;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 RRKRSRKS 135
|||||
Db 19 RRKRSRKS 26

RESULT 11
S73210
ribosomal protein S12, chloroplast - red alga (Porphyra purpurea) chloroplast
C:Species: chloroplast Porphyra purpurea
C:Date: 19-Mar-1997 #sequence_revision 09-May-1997 #text_change 21-Jan-2000
C:Accession: S73210
R:Reith, M.; Munholland, J.

Plant Mol. Biol. Rep. 13, 333-335, 1995
A:Title: Complete nucleotide sequence of the Porphyra purpurea chloroplast genome.
A:Reference number: S73108
A:Accession: S73210
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-124 <REI>
A:Cross-references: EMBL:U38804; NID:g1276652; PIDN:AAC08175.1; PID:g1276755
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1995
C:Genetics:
A:Gene: rps12
A:Genome: chloroplast
C:Superfamily: Escherichia coli ribosomal protein S12
C:Keywords: chloroplast; protein biosynthesis; ribosome
F:89/Modified site: beta-methylthiospartic acid (Asp) #status predicted

Query Match 3.6%; Score 7; DB 2; Length 124;
Best Local Similarity 100.0%; Pred. No. 9.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 RRKRSRKS 134
|||||
Db 110 RRKRSRKS 116

RESULT 12
F89858
conserved hypothetical protein SA0786 [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 17-May-2002
C:Accession: F89858
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; O
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: F89858
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-127 <KUR>
A:Cross-references: GB:BA000018; PID:g13700728; PIDN:BA842025.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: SA0786
C:Superfamily: Bacillus subtilis hypothetical protein yutD

Query Match 3.6%; Score 7; DB 2; Length 127;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 EYCNFGC 105
|||||
Db 75 EYCNFGC 81

RESULT 13
E70416
hypothetical protein aq_1343 - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 05-Nov-1999
C:Accession: E70416
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;
V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666; PMID:9537320
A:Accession: E70416
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-160 <AQF>
A:Cross-references: GB:AE000735; NID:g2983749; PIDN:AAC07319.1; PID:g2983759; GB:AE00
A:Experimental source: strain VF5

A;Residues: 1-91 <GLA>
 A;Cross-references: GB:NC_003210; PIDN:CAD00480.1; PID:gl6411890; GSPDB:GN00177
 A;Experimental source: strain EGD-e
 C;Genetics:
 A;Gene: lmo2402
 C;Superfamily: Bacillus subtilis hypothetical protein yutD

Query Match 5.7%; Score 11; DB 2; Length 91;
 Best Local Similarity 100.0%; Pred. No. 0.00035;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 99 EYCNFGCAYFV 109
 |||||
 Db 76 EYCNFGCAYFV 86

RESULT 5

AH1744
 B; subtilis YutD protein homolog lin2501 [imported] - Listeria innocua (strain Clip11262)
 C;Species: Listeria innocua
 C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 17-May-2002
 C;Accession: AH1744
 R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker, J.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; D.; Jones, L.M.; Karst, U.
 Science 294, 849-852, 2001
 A;Authors: Kref, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mak, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of Listeria species.
 A;Reference number: AB1077; MUID:21537279; PMID:11679669
 A;Accession: AH1744
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-91 <GLA>
 A;Cross-references: GB:AL592022; PIDN:CAC97728.1; PID:gl6415023; GSPDB:GN00178
 A;Experimental source: strain Clip11262
 C;Genetics:
 A;Gene: lin2501
 C;Superfamily: Bacillus subtilis hypothetical protein yutD

Query Match 5.7%; Score 11; DB 2; Length 91;
 Best Local Similarity 100.0%; Pred. No. 0.00035;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 99 EYCNFGCAYFV 109
 |||||
 Db 76 EYCNFGCAYFV 86

RESULT 6

A84079
 hypothetical protein BH3433 [imported] - Bacillus halodurans (strain C-125)
 C;Species: Bacillus halodurans
 C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 17-May-2002
 C;Accession: A84079
 R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira, Nucleic Acids Res. 28, 4317-4331, 2000
 A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
 A;Reference number: A83650; MUID:20512582; PMID:11058132
 A;Accession: A84079
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-93 <STO>
 A;Cross-references: GB:AP001518; GB:BA000004; MID:gl0175792; PIDN:BA07152.1; GSPDB:GN00179
 A;Experimental source: strain C-125
 C;Genetics:
 A;Gene: BH3433
 C;Superfamily: Bacillus subtilis hypothetical protein yutD

Query Match 5.7%; Score 11; DB 2; Length 93;
 Best Local Similarity 100.0%; Pred. No. 0.00035;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 99 EYCNFGCAYFV 109
 |||||
 Db 75 EYCNFGCAYFV 85

RESULT 7

T23176
 hypothetical protein K01D12.9 - Caenorhabditis elegans
 C;Species: Caenorhabditis elegans
 C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 11-Jan-2000
 C;Accession: T23176
 R;Dobson, R.
 submitted to the EMBL Data Library, June 1996
 A;Reference number: Z19703
 A;Accession: T23176
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-109 <WIL>
 A;Cross-references: EMBL:Z75543; PIDN:CAA99865.1; GSPDB:GN00023; CESP:K01D12.9
 A;Experimental source: clone K01D12
 C;Genetics:
 A;Gene: CESP:K01D12.9
 A;Map position: 5
 A;Introns: 38/1
 C;Superfamily: hypothetical protein K01D12.8

Query Match 4.2%; Score 8; DB 2; Length 109;
 Best Local Similarity 100.0%; Pred. No. 0.73;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 157 SKRKRDYK 164
 |||||
 Db 73 SKRKRDYK 80

RESULT 8

B86647
 hypothetical protein yb1c [imported] - Lactococcus lactis subsp. lactis (strain IL140)
 C;Species: Lactococcus lactis subsp. lactis
 C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
 C;Accession: B86647
 R;Bolotin, A.; Wincker, P.; Manger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; EH Genome Res. 11, 731-753, 2001
 A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis
 A;Reference number: A86625; MUID:21235186; PMID:11337471
 A;Accession: B86647
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-251 <STO>
 A;Cross-references: GB:AE005176; PID:gl2723030; PIDN:AAK04276.1; GSPDB:GN00146
 A;Experimental source: strain IL1403
 C;Genetics:
 A;Gene: yb1c

Query Match 4.2%; Score 8; DB 2; Length 251;
 Best Local Similarity 100.0%; Pred. No. 1.5;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 LYNKYKPG 17
 |||||
 Db 10 LYNKYKPG 17

RESULT 9

S64928
 hypothetical protein YLR094c - yeast (Saccharomyces cerevisiae)
 N;Alternate names: hypothetical protein L2534
 C;Species: Saccharomyces cerevisiae
 C;Date: 01-Aug-1995 #sequence_revision 24-May-1996 #text_change 19-Apr-2002
 C;Accession: S64928
 R;Benes, V.; Rechmann, S.; Nentwich, U.; Schwager, C.; Ansorge, W.; Voss, H.
 submitted to the Protein Sequence Database, May 1996
 A;Reference number: S64920

979 5 2.6 119 2 T04243 acidic ribosomal p
 980 5 2.6 120 2 I34487 T-cell receptor al
 981 5 2.6 120 2 AG0808 conserved hypothet
 982 5 2.6 120 2 T44670 V-ATPase proteolip
 983 5 2.6 120 2 F70966 hypothetical prote
 984 5 2.6 120 2 S73654 hypothetical prote
 985 5 2.6 120 2 C82909 hypothetical prote
 986 5 2.6 120 2 H33761 hypothetical prote
 987 5 2.6 120 2 T41017 hypothetical HXC-2
 988 5 2.6 121 1 J01646 nucleic acid-bind
 989 5 2.6 121 1 H71053 hypothetical prote
 990 5 2.6 121 2 I27887 Ig heavy chain v r
 991 5 2.6 121 2 A27888 Ig heavy chain v r
 992 5 2.6 121 2 H27887 Ig heavy chain v r
 993 5 2.6 121 2 A03327 Ig kappa chain - h
 994 5 2.6 121 2 A70443 ribosomal protein
 995 5 2.6 121 2 D75089 hypothetical prote
 996 5 2.6 121 2 S48420 probable membrane
 997 5 2.6 121 2 C72627 hypothetical prote
 998 5 2.6 121 2 AF0107 hypothetical prote
 999 5 2.6 122 1 R5DV14 ribosomal protein
 1000 5 2.6 122 2 B30515 Ig heavy chain v r

ALIGNMENTS

RESULT 1
 F70023
 Hypothetical protein yutD - Bacillus subtilis
 C:Species: Bacillus subtilis
 C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
 C:Accession: F70023
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte
 C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
 A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizzi, A.; Gall
 iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
 Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinol
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue
 Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetel
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Satoh, T.; Scanlon
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron
 akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpsira, P.; Tognoni, A.; Tosato, V.; Uchiyama
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
 A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
 A:Reference number: A69580; MUID:98044033; PMID:9384377
 A:Accession: F70023
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-102 <KUN>
 A:Cross-references: GB:Z99120; GB:AL009126; NID:g2635613; PIDN:CAB15221.1; PID:g2635728
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: yutD
 C:Superfamily: Bacillus subtilis hypothetical protein yutD

Query Match 6.2%; Score 12; DB 2; Length 102;
 Best Local Similarity 100.0%; Pred. No. 3.2e-05;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 99 EYCNFGCAVFL 110
 |||||
 DB 86 EYCNFGCAVFL 97

RESULT 2
 H95088
 conserved hypothetical protein SP0767 [imported] - Streptococcus pneumoniae (strain TIGR
 C:Species: Streptococcus pneumoniae
 C>Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001

C:Accession: H95088
 R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H
 on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzappl
 nson, T.; Hickey, E.K.; Holt, I.E.
 Science 293, 498-506, 2001
 A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris
 A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
 A:Reference number: A95000; MUID:21357209; PMID:11463916
 A:Accession: H95088
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-170 <KUR>
 A:Cross-references: GB:AF005672; PIDN:AAK74905.1; PID:gl4972242; GSPDB:GN00164; TIGR:
 A:Experimental source: strain TIGR4
 C:Genetics:
 A:Gene: SP0767

Query Match 6.2%; Score 12; DB 2; Length 170;
 Best Local Similarity 100.0%; Pred. No. 4.9e-05;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 69 NEQLRLRGFYKD 80
 |||||
 DB 69 NEQLRLRGFYKD 80

RESULT 3
 C97956
 Hypothetical protein spr0675 [imported] - Streptococcus pneumoniae (strain R6)
 C:Species: Streptococcus pneumoniae
 C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
 C:Accession: C97956
 R:Hoskins, J.A.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.
 e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.
 Y, P.; Sun, P.M.; Winkler, M.E.
 J. Bacteriol. 183, 5709-5717, 2001
 A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.
 A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
 A:Reference number: A37872; MUID:21429245; PMID:11544234
 A:Accession: C97956
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-176 <KUR>
 A:Cross-references: GB:AE007317; PIDN:AAK99479.1; PID:gl5458263; GSPDB:GN00174
 C:Genetics:
 A:Gene: spr0675

Query Match 6.2%; Score 12; DB 2; Length 176;
 Best Local Similarity 100.0%; Pred. No. 5e-05;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 69 NEQLRLRGFYKD 80
 |||||
 DB 69 NEQLRLRGFYKD 80

RESULT 4
 AB1375
 B. subtilis yutD protein homolog lmo2402 [imported] - Listeria monocytogenes (strain
 C:Species: Listeria monocytogenes
 C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 17-May-2002
 C:Accession: AB1375
 R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec
 .; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi,
 D.; Jones, L.M.; Karst, U.
 Science 294, 849-852, 2001
 A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkut, G.; Madseno, E.; Maitournam, A.;
 Ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla
 A:Title: Comparative genomics of Listeria species.
 A:Reference number: AB1077; MUID:21537279; PMID:11679669
 A:Accession: AB1375
 A:Status: preliminary
 A:Molecule type: DNA

833	5	2.6	96	2	AF0741	conserved hypot	906	108	2	AF1935	hypothetical prote
834	5	2.6	96	2	B85797	yebG protein [simi	907	109	1	K3HUS1	Ig kappa chain V-I
835	5	2.6	96	2	F90948	yebG protein [simi	908	109	1	K3HUGO	Ig kappa chain V-I
836	5	2.6	96	2	S76170	hypothetical prote	909	109	1	K3HUTI	Ig kappa chain V-I
837	5	2.6	96	2	C95403	hypothetical prote	910	109	1	K3HOML	Ig kappa chain V-I
838	5	2.6	96	2	AF0151	probable bacteriop	911	109	2	PH0963	Ig kappa chain V-I
839	5	2.6	96	2	AE3640	hypothetical prote	912	109	2	G30601	Ig kappa chain V-I
840	5	2.6	97	2	AC3259	LSU ribosomal prot	913	109	2	G30601	Ig kappa chain V-I
841	5	2.6	98	1	OPCH2	acylphosphatase (E	914	109	2	A30608	Ig kappa chain V-I
842	5	2.6	98	2	G81251	NADH2 dehydrogenas	915	109	2	D30601	Ig kappa chain V-I
843	5	2.6	98	2	S40361	Ig kappa chain V-J	916	109	2	F30607	Ig kappa chain V-I
844	5	2.6	98	2	D72800	gpA protein - Myco	917	109	2	G30601	Ig kappa chain V-I
845	5	2.6	98	2	A46026	group I intron end	918	109	2	G30607	Ig kappa chain V-I
846	5	2.6	99	2	JH0429	Ig kappa chain V-I	919	109	2	H30601	Ig kappa chain V-I
847	5	2.6	99	2	S12161	polyubiquitin - la	920	109	2	S57098	probable membrane
848	5	2.6	99	2	S62880	ubiquitin/ribosoma	921	109	2	C82913	conserved hypot
849	5	2.6	99	2	H90594	hypothetical prote	922	109	2	H81342	hypothetical prote
850	5	2.6	100	1	K3HUNG	Ig kappa chain pre	923	110	2	S20635	Ig kappa chain V r
851	5	2.6	100	2	T08248	TATA-binding trans	924	110	2	E30607	Ig kappa chain V-I
852	5	2.6	100	2	D71882	hypothetical prote	925	110	2	S44120	Ig kappa chain V-J
853	5	2.6	100	2	G71549	hypothetical prote	926	110	2	A24092	T-cell receptor al
854	5	2.6	100	2	F84766	hypothetical prote	927	110	2	B24092	T-cell receptor al
855	5	2.6	100	2	AC0684	hypothetical prote	928	110	2	S15608	hypothetical prote
856	5	2.6	100	2	AE3219	hypothetical prote	929	110	2	B70343	conserved hypot
857	5	2.6	100	2	S41505	12RNP2 protein - S	930	110	2	JC2327	hypothetical 12.6K
858	5	2.6	100	2	T50611	hypothetical prote	931	110	2	F69760	hypothetical prote
859	5	2.6	100	2	G96948	uncharacterized sm	932	110	2	S15153	hypothetical prote
860	5	2.6	101	2	T43777	ribosomal protein	933	110	2	B70601	hypothetical prote
861	5	2.6	101	2	H84076	hypothetical prote	934	110	2	B97412	hypothetical prote
862	5	2.6	101	2	C72741	hypothetical prote	935	111	2	S40359	Ig kappa chain V-J
863	5	2.6	101	2	A05124	hypothetical prote	936	111	2	E75033	hypothetical prote
864	5	2.6	101	2	D72328	hypothetical prote	937	112	2	A86604	periplasmic divale
865	5	2.6	101	2	S65494	sperm-specific pro	938	112	2	B72019	periplasmic divale
866	5	2.6	101	2	C97704	hypothetical prote	939	112	2	S74214	fmdB protein - Met
867	5	2.6	101	2	AH2395	hypothetical prote	940	112	2	T07633	polyubiquitin 6 -
868	5	2.6	102	2	S01033	hypothetical prote	941	113	2	S26266	T-cell receptor be
869	5	2.6	102	2	S61990	hypothetical prote	942	113	2	PT0741	T-cell receptor be
870	5	2.6	102	2	S65340	hypothetical prote	943	113	2	S17385	T-cell receptor be
871	5	2.6	102	2	T29651	hypothetical prote	944	113	2	S17386	T-cell receptor be
872	5	2.6	103	2	S28823	transcription fact	945	113	2	S26262	T-cell receptor be
873	5	2.6	103	2	PH1045	Ig light chain V r	946	113	2	S26263	T-cell receptor be
874	5	2.6	103	2	B81160	conserved hypot	947	113	2	T44042	chemokine [impor
875	5	2.6	103	2	D52350	conserved hypot	948	113	2	C89828	conserved hypot
876	5	2.6	103	2	G86346	F24J8.2 protein -	949	114	2	S46375	Ig kappa chain V-J
877	5	2.6	104	2	A24424	T-cell receptor be	950	114	2	F83959	ribosomal protein
878	5	2.6	104	2	PH0964	Ig kappa chain V r	951	114	2	B24733	myosin heavy chain
879	5	2.6	104	2	S28696	hypothetical prote	952	114	2	G72611	hypothetical prote
880	5	2.6	105	1	B71334	probable RNA-bind	953	114	2	T09490	hypothetical prote
881	5	2.6	105	2	S11296	polyubiquitin (clo	954	114	2	T17699	arginine-rich prot
882	5	2.6	105	2	T14406	pectin esterase ho	955	114	2	B97036	probable metal-bin
883	5	2.6	105	2	AH2094	RNA-binding protei	956	114	2	C97205	uncharacterized pr
884	5	2.6	105	2	S27493	nodC protein - Bra	957	115	2	JC5908	T cell receptor va
885	5	2.6	105	2	S56482	hypothetical prote	958	115	2	H75543	hypothetical prote
886	5	2.6	105	2	A48943	phage antigenic de	959	115	2	B90263	hypothetical prote
887	5	2.6	106	2	S71947	ubiquitin thiolest	960	116	2	B25521	Ig kappa chain pre
888	5	2.6	106	2	T25815	hypothetical prote	961	116	2	B27594	Ig kappa chain pre
889	5	2.6	106	2	D89551	protein K10C2.6 [i	962	116	2	C27594	Ig kappa chain pre
890	5	2.6	106	2	D97815	hypothetical prote	963	116	2	S62908	ubiquitin precurs
891	5	2.6	106	2	H89982	truncated map-w pr	964	116	2	S24989	ribosomal protein
892	5	2.6	107	1	ERAD51	early E3A 12.5K pr	965	116	2	S50805	hypothetical prote
893	5	2.6	107	2	PH0965	Ig kappa chain V r	966	116	2	S64326	hypothetical prote
894	5	2.6	107	2	S07663	hypothetical prote	967	116	2	C72265	hypothetical prote
895	5	2.6	107	2	S07818	homeotic protein H	968	117	2	I68824	T-cell receptor al
896	5	2.6	108	2	C30608	Ig kappa chain V-I	969	117	2	T24263	hypothetical prote
897	5	2.6	108	2	S33988	Ig kappa chain V r	970	118	2	T03036	Ig light chain - h
898	5	2.6	108	2	E30609	Ig kappa chain V-I	971	118	2	S25001	ubiquitin/ribosoma
899	5	2.6	108	2	A44151	Ig kappa chain V r	972	118	2	C41868	traE1 protein - En
900	5	2.6	108	2	H72224	hypothetical prote	973	118	2	S62422	hypothetical prote
901	5	2.6	108	2	T38079	very hypot	974	118	2	S75222	hypothetical prote
902	5	2.6	108	2	S44804	F10E9.5 protein -	975	119	2	E89999	holo-ACP synthase
903	5	2.6	108	2	T49144	hypothetical prote	976	119	2	T07487	hypothetical prote
904	5	2.6	108	2	S74608	hypothetical prote	977	119	2	D83723	hypothetical prote
905	5	2.6	108	2	G69261	hypothetical prote	978	119	2	E81655	conserved hypot

687	5	2.6	64	2	S58141	gene 12 protein -	760	5	2.6	84	2	G64231	modulation protein
688	5	2.6	64	2	T12921	hypothetical prote	761	5	2.6	84	2	S36762	dephosphin long fo
689	5	2.6	64	2	A12881	hypothetical prote	762	5	2.6	84	2	H59106	hypothetical prote
690	5	2.6	65	2	G02999	BRCA1 - rhesus mac	763	5	2.6	84	2	E97807	hypothetical prote
691	5	2.6	66	2	B82818	hypothetical prote	764	5	2.6	85	2	A60622	probable bacteriop
692	5	2.6	67	2	T12871	hypothetical prote	765	5	2.6	85	2	AD2497	hypothetical prote
693	5	2.6	67	2	F97130	probable transcrip	766	5	2.6	86	2	T12077	late embryogenesis
694	5	2.6	67	2	AF2126	hypothetical prote	767	5	2.6	86	2	T12077	E4 protein - human
695	5	2.6	68	2	AG3217	hypothetical prote	768	5	2.6	87	1	W4WL51	hypothetical prote
696	5	2.6	68	2	S01884	ubiquitin / riboso	769	5	2.6	87	2	S16843	Ig kappa chain V r
697	5	2.6	69	2	A64527	hypothetical prote	770	5	2.6	87	2	S16842	Ig kappa chain V r
698	5	2.6	69	2	C71980	hypothetical prote	771	5	2.6	87	2	S16841	Ig kappa chain V r
699	5	2.6	69	2	T03361	gene e6 protein -	772	5	2.6	87	2	S16823	CP75 protein - pha
700	5	2.6	70	2	AD4841	dynammin - bovine (773	5	2.6	87	2	S09534	hypothetical prote
701	5	2.6	70	2	AD2729	conserved hypothet	774	5	2.6	87	2	E97510	ubiquitin precursor
702	5	2.6	71	1	S15869	ribosomal protein	775	5	2.6	88	1	UQSY	repressor protein
703	5	2.6	71	1	B34209	trichohyalin - gui	776	5	2.6	88	2	T13183	9.8K protein - Ber
704	5	2.6	71	2	F96600	protein F14J16.22	777	5	2.6	88	2	B36113	hypothetical prote
705	5	2.6	71	2	A81037	probable secretion	778	5	2.6	88	2	T34429	Ig kappa chain V r
706	5	2.6	71	2	A09927	probable secretory	779	5	2.6	89	2	S34096	lipopolysaccharide
707	5	2.6	72	2	S34404	ribosomal protein	780	5	2.6	89	2	T35348	probable membrane
708	5	2.6	72	2	A81223	hypothetical prote	781	5	2.6	89	2	S74690	conserved hypothet
709	5	2.6	72	2	T31010	hypothetical prote	782	5	2.6	90	1	S78047	DNA-directed RNA p
710	5	2.6	73	2	S34406	ribosomal protein	783	5	2.6	90	2	E86681	prophage pil prote
711	5	2.6	73	2	A83195	conserved hypothet	784	5	2.6	90	2	T17573	hypothetical prote
712	5	2.6	73	2	T25926	hypothetical prote	785	5	2.6	90	2	T29013	hypothetical prote
713	5	2.6	74	2	E84234	hypothetical prote	786	5	2.6	90	2	A97755	hypothetical prote
714	5	2.6	76	1	UQHU	ubiquitin - human	787	5	2.6	91	2	S67940	Ig kappa chain V r
715	5	2.6	76	1	UQBO	ubiquitin - bovine	788	5	2.6	91	2	S67940	Ig kappa chain V r
716	5	2.6	76	1	UQOA	ubiquitin - oat	789	5	2.6	91	2	E69087	signal recognition
717	5	2.6	76	1	UQUVSF	ubiquitin - fall a	790	5	2.6	91	2	B72200	hypothetical prote
718	5	2.6	76	1	A70028	hypothetical prote	791	5	2.6	92	2	S37524	Ig kappa chain V r
719	5	2.6	76	2	A26087	ubiquitin - fruit	792	5	2.6	92	2	S37513	Ig kappa chain V r
720	5	2.6	76	2	S28203	ubiquitin - rabbit	793	5	2.6	92	2	S37517	Ig kappa chain V r
721	5	2.6	76	2	S42750	polyubiquitin UB2	794	5	2.6	92	2	S37507	Ig kappa chain V r
722	5	2.6	76	2	S12114	polyubiquitin - in	795	5	2.6	92	2	S37502	Ig kappa chain V r
723	5	2.6	76	2	S31653	ubiquitin - Giardia	796	5	2.6	92	2	S37514	Ig kappa chain V r
724	5	2.6	76	2	D69905	hypothetical prote	797	5	2.6	92	2	S37503	Ig kappa chain V r
725	5	2.6	76	2	H97842	hypothetical prote	798	5	2.6	92	2	S37518	Ig kappa chain V r
726	5	2.6	76	2	A11304	hypothetical prote	799	5	2.6	92	2	T29701	hypothetical prote
727	5	2.6	77	1	UQUT	ubiquitin - Trypan	800	5	2.6	92	2	B97044	hypothetical prote
728	5	2.6	77	1	UQNVAC	ubiquitin - Autogr	801	5	2.6	92	2	C89812	hypothetical prote
729	5	2.6	77	2	S29238	ubiquitin - Entamo	802	5	2.6	93	1	R38219	ribosomal protein
730	5	2.6	77	2	S34334	ubiquitin - Leishm	803	5	2.6	93	1	R32M19	ribosomal protein
731	5	2.6	77	2	S19799	ubiquitin - potato	804	5	2.6	93	2	S37508	Ig kappa chain V r
732	5	2.6	77	2	S16263	ubiquitin precursor	805	5	2.6	93	2	S37526	Ig kappa chain V r
733	5	2.6	77	2	C72854	viral ubiquitin -	806	5	2.6	93	2	S37528	Ig kappa chain V r
734	5	2.6	77	2	T41781	ubiquitin Orf35 -	807	5	2.6	93	2	S37505	Ig kappa chain V r
735	5	2.6	77	2	B90849	hypothetical prote	808	5	2.6	93	2	T10294	ubiquitin-like pro
736	5	2.6	77	2	H85706	unknown protein en	809	5	2.6	93	2	JQ2029	ubiquitin - Orgyia
737	5	2.6	78	1	UQUTC	ubiquitin precursor	810	5	2.6	93	2	JN0301	ribosomal protein
738	5	2.6	78	1	G69508	conserved hypothet	811	5	2.6	93	2	S31184	outer membrane pro
739	5	2.6	78	2	A47243	T-cell receptor be	812	5	2.6	93	2	C81783	hypothetical prote
740	5	2.6	78	2	C49039	T-cell receptor be	813	5	2.6	93	2	A69982	hypothetical prote
741	5	2.6	78	2	T03930	gene GUT15 protein	814	5	2.6	93	2	A33472	transcription regu
742	5	2.6	78	2	B97027	feoA-like protein,	815	5	2.6	93	2	S72863	hypothetical prote
743	5	2.6	79	2	D86945	hypothetical prote	816	5	2.6	93	2	E69163	hypothetical prote
744	5	2.6	79	2	T44149	hypothetical prote	817	5	2.6	94	2	S11248	ubiquitin / riboso
745	5	2.6	79	2	B83400	hypothetical prote	818	5	2.6	94	2	D84117	transcription regu
746	5	2.6	79	2	B69874	hypothetical prote	819	5	2.6	95	2	JT0492	ubiquitin 2 - Tetr
747	5	2.6	80	2	S35775	T-cell receptor al	820	5	2.6	95	2	S25154	ubiquitin / riboso
748	5	2.6	80	2	T28305	ORF MSV144 probabl	821	5	2.6	95	2	T41855	AcMNPV orf117 - Bo
749	5	2.6	80	2	E83128	hypothetical prote	822	5	2.6	95	2	G70244	conserved hypothet
750	5	2.6	80	2	I68668	transcription fact	823	5	2.6	95	2	D83360	hypothetical prote
751	5	2.6	80	2	B95851	hypothetical prote	824	5	2.6	95	2	H83237	hypothetical prote
752	5	2.6	82	2	H81138	conserved hypothet	825	5	2.6	95	2	F81873	probable membrane
753	5	2.6	82	2	H82208	hypothetical prote	826	5	2.6	95	2	D64389	hypothetical prote
754	5	2.6	82	2	D64329	hypothetical prote	827	5	2.6	95	2	T40664	very hypothetical
755	5	2.6	82	2	E97781	RP511 protein homo	828	5	2.6	95	2	T13135	protein gp48 - pha
756	5	2.6	83	2	S75314	phycocyanin-associ	829	5	2.6	96	1	ASLJSC	vpr protein - siml
757	5	2.6	83	2	T07252	hypothetical prote	830	5	2.6	96	2	A30601	Ig kappa chain V-I
758	5	2.6	83	2	H89471	protein B0294.2 [i	831	5	2.6	96	2	H70340	hypothetical prote
759	5	2.6	84	2	S01981	ribosomal protein	832	5	2.6	96	2	H64946	yebG protein - Esc

541	6	3.1	983	2	B83055	probable two-compo	614	2052	2	T18290	FYVE finger-contai
542	6	3.1	994	2	B82843	valyl-tRNA synthet	615	2094	2	S33124	tptr protein - huma
543	6	3.1	994	2	S00960	hypothetical prote	616	2172	2	T20145	hypothetical prote
544	6	3.1	1000	2	JE0110	mitotic control pr	617	2364	1	A4159	spectrin beta-G ch
545	6	3.1	1002	2	T43236	carboxypeptidase C	618	2388	2	JE0271	beta spectrin, bet
546	6	3.1	1002	2	T02488	hypothetical prote	619	2458	2	T17420	probable polyketid
547	6	3.1	1007	2	T32285	hypothetical prote	620	2774	2	A43359	microtubule-associ
548	6	3.1	1011	2	S11177	vacuolar sorting p	621	2897	2	B48666	cell proliferation
549	6	3.1	1011	2	T20785	hypothetical prote	622	3255	2	A48666	cell proliferation
550	6	3.1	1032	2	A11697	alpha-mannosidase	623	3295	2	AE0074	probable adhesin Y
551	6	3.1	1033	2	S19247	cell adhesion prot	624	6658	2	T13931	projectin - fruit
552	6	3.1	1036	2	AG1326	alpha-mannosidase	625	7576	2	T17428	FK506 polyketide s
553	6	3.1	1039	2	S18199	myosin heavy chain	626	8243	2	T31307	type I fatty acid
554	6	3.1	1053	2	S34172	sucrose-phosphate	627	15	2	S10891	ubiquitin thiolest
555	6	3.1	1057	1	OYRTR	atrial natriuretic	628	15	2	PT0093	ubiquitin thiolest
556	6	3.1	1057	1	OYMSAR	atrial natriuretic	629	20	2	S28405	ubiquitin thiolest
557	6	3.1	1057	2	I57963	natriuretic peptid	630	20	2	S27350	lamin B receptor -
558	6	3.1	1057	2	I55319	guanylyl cyclase A	631	22	2	S19802	lysophospholipase
559	6	3.1	1061	1	OYH0AR	natriuretic peptid	632	23	2	S45031	ubiquitin - potato
560	6	3.1	1072	2	T18802	hypothetical prote	633	25	2	S17561	homeotic protein S
561	6	3.1	1073	2	D87596	TonB-dependent rec	634	27	2	B47295	ubiquitin thiolest
562	6	3.1	1084	2	S57681	CRM1 protein - yea	635	30	2	PD0006	homeodomain protei
563	6	3.1	1091	2	S57112	JSN1 protein - yea	636	34	2	H30607	cysteine synthase
564	6	3.1	1113	2	S28925	nuclear pore compl	637	34	2	J21080	Ig kappa chain V-I
565	6	3.1	1118	2	T27865	hypothetical prote	638	36	2	SC2006	sperm protein EM1
566	6	3.1	1159	2	A34308	Ca2+-transporting	639	36	2	A69827	differentiation in
567	6	3.1	1160	2	T31688	Ca2+-transporting	640	38	2	E49164	hypothetical prote
568	6	3.1	1163	2	E71481	probable snf/snf h	641	38	2	C83729	chromogranin-B - r
569	6	3.1	1166	2	G86596	SWI/SNF family hel	642	41	2	A45942	hypothetical prote
570	6	3.1	1166	2	C72027	swi/snf family hel	643	41	2	T45906	myosin heavy chain
571	6	3.1	1175	2	H83437	hypothetical prote	644	42	2	S20863	hypothetical prote
572	6	3.1	1188	2	G72734	hypothetical prote	645	42	2	JH0383	ubiquitin - Norway
573	6	3.1	1210	2	I39410	AF-4 protein, spli	646	42	2	JH0384	hypothetical 5.1K
574	6	3.1	1213	2	A58198	serine/proline-ric	647	42	2	I65249	hypothetical 5.1K
575	6	3.1	1217	2	T00607	hypothetical prote	648	42	2	T07324	homeoprotein 13A -
576	6	3.1	1230	2	T22458	hypothetical prote	649	44	2	H90641	hypothetical prote
577	6	3.1	1272	2	C90593	hypothetical prote	650	44	2	H85492	hypothetical prote
578	6	3.1	1278	2	B70236	hypothetical prote	651	45	2	C82451	hypothetical prote
579	6	3.1	1333	2	B84601	probable retroelem	652	46	2	PD0024	protein kinase (EC
580	6	3.1	1335	2	H75511	DNA polymerase III	653	50	2	B64035	hypothetical prote
581	6	3.1	1338	2	T02206	hypothetical prote	654	51	2	S26101	outer membrane pro
582	6	3.1	1344	2	T42637	hypothetical prote	655	51	2	T07326	hypothetical prote
583	6	3.1	1366	2	T35985	probable large pro	656	53	2	A56512	hypothetical prote
584	6	3.1	1403	2	S64142	hypothetical prote	657	54	2	S40381	nuclear pore prote
585	6	3.1	1404	2	T19277	hypothetical prote	658	54	2	A25521	Ig kappa chain V-I
586	6	3.1	1407	1	S28589	trichohyalin - rab	659	55	1	R5NT32	Ig kappa chain V r
587	6	3.1	1444	2	T18956	angiogenesis inhib	660	55	2	S43121	ribosomal protein
588	6	3.1	1455	2	A70199	hypothetical prote	661	55	2	T06496	ubiquitin - basidi
589	6	3.1	1477	2	S64616	YOR1 protein - yea	662	55	2	A28911	ubiquitin/ribosoma
590	6	3.1	1489	2	S60416	DNA helicase YGL15	663	55	2	T09464	male accessory gla
591	6	3.1	1491	1	S43793	copper-transportin	664	55	2	E48362	H+-transporting tw
592	6	3.1	1496	2	I51170	carbamyl phosphate	665	56	2	S45027	hypothetical prote
593	6	3.1	1502	2	T48309	hypothetical prote	666	56	2	G90278	sox1 protein - Her
594	6	3.1	1512	2	G85090	hypothetical prote	667	56	2	C82226	hypothetical prote
595	6	3.1	1519	2	T13402	DNA (cytosine-5)-	668	56	2	AC1987	hypothetical prote
596	6	3.1	1547	2	T13847	sno protein - frui	669	57	2	G84072	hypothetical prote
597	6	3.1	1549	1	A40691	trichohyalin - she	670	57	2	G97083	hypothetical prote
598	6	3.1	1564	2	S55517	probable transport	671	58	2	B48663	pyruvate kinase (E
599	6	3.1	1592	2	T16055	hypothetical prote	672	59	2	D81084	hypothetical prote
600	6	3.1	1601	2	AE2011	hypothetical prote	673	60	2	S35929	hypothetical prote
601	6	3.1	1605	2	T31435	DNA-directed RNA p	674	60	2	JT0491	pyruvate kinase (E
602	6	3.1	1653	2	A36349	clathrin heavy cha	675	60	2	S09227	ubiquitin 1 - Tetr
603	6	3.1	1696	2	T27447	hypothetical prote	676	60	2	B69189	homeotic protein H
604	6	3.1	1726	2	A39401	merozoite surface	677	62	2	AH1301	conserved hypothet
605	6	3.1	1731	2	A45604	major blood-stage	678	62	2	TH1673	ribosomal protein
606	6	3.1	1755	2	F82618	chemotaxis-related	679	62	2	T10302	ribosomal protein
607	6	3.1	1778	2	T50074	probable nucleopor	680	62	2	C97621	hypothetical prote
608	6	3.1	1822	2	S63985	collagen alpha 2 c	681	62	2	AI2130	hypothetical prote
609	6	3.1	1852	1	VJCH2	vitellogenin II pr	682	63	2	B30606	T-cell receptor al
610	6	3.1	1898	1	A45973	trichohyalin - hum	683	63	2	I48071	GTP binding protei
611	6	3.1	1906	2	AD2443	hypothetical prote	684	63	2	G96517	protein T2E6.24 [l
612	6	3.1	1939	2	I48175	myosin heavy chain	685	63	2	F84501	10-formyltetrahydr
613	6	3.1	1944	2	T40065	tRNA-splicing endo	686	64	2	S17384	T-cell receptor be

395	6	3.1	550	1	E28544	methyl coenzyme M	468	6	3.1	728	2	D86278	hypothetical prote
396	6	3.1	550	2	B69022	methyl coenzyme M	469	6	3.1	728	2	A54603	transcription fact
397	6	3.1	552	2	C96563	probable protein k	470	6	3.1	732	2	T18567	hypothetical prote
398	6	3.1	553	2	T14354	probable somatic e	471	6	3.1	749	2	T34090	hypothetical prote
399	6	3.1	554	2	T31783	hypothetical prote	472	6	3.1	752	2	D95246	glycogen phospho
400	6	3.1	557	2	A56690	esterase - Caenorh	473	6	3.1	752	2	A98111	maltoextrin phosp
401	6	3.1	557	2	E72282	oligopeptide ABC t	474	6	3.1	752	2	T16508	hypothetical prote
402	6	3.1	560	2	S58741	probable site-spec	475	6	3.1	752	2	T26508	hypothetical prote
403	6	3.1	560	2	T40608	conserved hypothet	476	6	3.1	754	1	JC5314	CDC38/cdc2-like Ki
404	6	3.1	562	2	T42250	polypeptide N-acet	477	6	3.1	754	1	AF2849	methyl-accepting c
405	6	3.1	562	2	T26242	hypothetical prote	478	6	3.1	761	2	T15776	hypothetical prote
406	6	3.1	563	2	T32504	hypothetical prote	479	6	3.1	765	2	T23582	hypothetical prote
407	6	3.1	564	2	T28949	hypothetical prote	480	6	3.1	768	2	T10796	3',5'-cyclic-nucle
408	6	3.1	566	2	T50536	succinate dehydrog	481	6	3.1	768	2	D97626	hypothetical prote
409	6	3.1	568	2	T26243	hypothetical prote	482	6	3.1	773	2	T00502	probable receptor-
410	6	3.1	568	2	S15008	hypothetical prote	483	6	3.1	779	2	E97778	endopeptidase La (
411	6	3.1	570	2	E29525	gene disco protein	484	6	3.1	784	2	AC0383	endopeptidase La (
412	6	3.1	570	2	G81380	probable AAA famil	485	6	3.1	790	2	S50082	nuclear cap bindin
413	6	3.1	571	2	F84718	hypothetical prote	486	6	3.1	791	2	E90566	conserved hypothet
414	6	3.1	573	2	E84687	hypothetical prote	487	6	3.1	796	2	E90566	hypothetical prote
415	6	3.1	573	2	H96744	probable cytosolic	488	6	3.1	803	2	T40514	Chaperonin hsp78p
416	6	3.1	574	2	T04249	hypothetical prote	489	6	3.1	807	2	T06409	valosin-containing
417	6	3.1	576	2	S22127	protein kinase pol	490	6	3.1	809	1	S60112	transitional endop
418	6	3.1	577	2	T04229	ABC-type transport	491	6	3.1	813	2	C97292	ATPases with chape
419	6	3.1	577	2	S47961	glucosyltransferas	492	6	3.1	815	2	T46169	CDC48-like protein
420	6	3.1	580	2	AH2138	ABC transporter Ar	493	6	3.1	815	2	B56708	extracellular sign
421	6	3.1	581	2	T05232	hypothetical prote	494	6	3.1	815	2	G97266	mannose-1-phosphat
422	6	3.1	583	2	T25690	hypothetical prote	495	6	3.1	817	2	H75035	probable membrane
423	6	3.1	585	2	C82157	hypothetical prote	496	6	3.1	822	2	S26872	DNA-directed RNA p
424	6	3.1	586	2	T49210	hypothetical prote	497	6	3.1	823	2	B87348	1,4-beta-D-glucan
425	6	3.1	588	2	E96533	probable Serine/Th	498	6	3.1	824	2	D64738	ATP-dependent hell
426	6	3.1	592	2	T48346	photoreceptor-inte	499	6	3.1	824	2	H90647	ATP-dependent hell
427	6	3.1	592	2	F86242	unknown protein, 9	500	6	3.1	824	2	H85498	helicase, ATP-depe
428	6	3.1	600	2	C83142	hypothetical prote	501	6	3.1	826	2	T02268	potassium transpor
429	6	3.1	601	2	H96740	hypothetical prote	502	6	3.1	826	2	T23072	hypothetical prote
430	6	3.1	603	2	B64444	hypothetical prote	503	6	3.1	827	2	F64607	DNA gyrase, sub A
431	6	3.1	607	2	B86453	F9L11.4 protein -	504	6	3.1	832	2	T33490	hypothetical prote
432	6	3.1	613	2	D86147	TiN6.3 protein - A	505	6	3.1	834	2	F83185	glycerol-3-phospha
433	6	3.1	617	2	T42249	polypeptide N-acet	506	6	3.1	835	1	S67669	transitional endop
434	6	3.1	618	2	T42248	polypeptide N-acet	507	6	3.1	843	2	T48355	transitional endop
435	6	3.1	619	2	F70480	conserved hypothet	508	6	3.1	847	2	G75270	cation-transportin
436	6	3.1	620	2	S48772	AFRI protein - yea	509	6	3.1	852	2	A85041	probable receptor
437	6	3.1	620	2	T11221	DNA polymerase hom	510	6	3.1	855	2	A81618	proliferating-cell
438	6	3.1	621	2	I57939	taurine transporte	511	6	3.1	859	1	S08418	3',5'-cyclic-GMP p
439	6	3.1	621	2	F70334	hypothetical prote	512	6	3.1	859	2	B34611	3',5'-cyclic-GMP p
440	6	3.1	627	2	AF1074	lytic murein trans	513	6	3.1	859	2	S13030	3',5'-cyclic-GMP p
441	6	3.1	628	2	D86466	69.4K hypothetical	514	6	3.1	859	2	S64952	hypothetical prote
442	6	3.1	629	2	T24469	hypothetical prote	515	6	3.1	862	2	T07775	lipoxigenase (EC 1
443	6	3.1	634	2	T27959	hypothetical prote	516	6	3.1	863	2	S37088	phycobilisome anch
444	6	3.1	639	2	AB3456	long-chain-fatty-a	517	6	3.1	863	2	T27958	hypothetical prote
445	6	3.1	639	2	B24810	beta-conglycinin a	518	6	3.1	864	2	S60441	hypothetical prote
446	6	3.1	639	2	A99225	mRNA 3'-end proces	519	6	3.1	870	2	T30110	hypothetical prote
447	6	3.1	640	2	C81108	ABC transporter, A	520	6	3.1	871	2	T43427	pob1 protein - fis
448	6	3.1	642	2	S31285	RNA polymerase II	521	6	3.1	875	2	S75377	probable helicase
449	6	3.1	647	2	T00370	hypothetical prote	522	6	3.1	881	2	C97362	protein sensor pro
450	6	3.1	649	2	G71283	probable ATP-depen	523	6	3.1	881	2	AD2580	two component sens
451	6	3.1	650	2	T21498	hypothetical prote	524	6	3.1	882	2	S23019	DNA-directed DNA p
452	6	3.1	653	2	S67035	probable membrane	525	6	3.1	882	2	F90201	DNA polymerase I (
453	6	3.1	655	2	A46270	Na and Cl dependen	526	6	3.1	892	2	T27005	autotaxin precuro
454	6	3.1	659	2	T41975	helicase-primate c	527	6	3.1	899	2	S49634	probable serine/th
455	6	3.1	662	2	E84725	ankyrin-like prote	528	6	3.1	900	2	F71426	hypothetical prote
456	6	3.1	677	2	T39713	zinc finger protel	529	6	3.1	904	2	T22457	hypothetical prote
457	6	3.1	689	2	S59797	hypothetical prote	530	6	3.1	906	1	S32607	trifunctional enzy
458	6	3.1	692	2	S48316	probable membrane	531	6	3.1	915	1	A55144	autotaxin precuro
459	6	3.1	693	2	B84647	hypothetical prote	532	6	3.1	923	2	T08033	probable serine/th
460	6	3.1	707	2	T29854	spectrin beta chai	533	6	3.1	924	1	JC1269	DNA-directed DNA p
461	6	3.1	708	2	T00205	hypothetical prote	534	6	3.1	925	2	C84538	probable LXR recep
462	6	3.1	710	2	S67098	probable membrane	535	6	3.1	927	2	T08034	probable serine/th
463	6	3.1	712	2	AG0419	ribonucleoside-tri	536	6	3.1	941	2	F71332	probable chromosom
464	6	3.1	713	2	T21201	hypothetical prote	537	6	3.1	950	2	G83167	valyl-tRNA synthet
465	6	3.1	714	2	T22454	hypothetical prote	538	6	3.1	960	1	A39651	discs-large tumor
466	6	3.1	722	2	AD1796	hypothetical membr	539	6	3.1	962	2	T22459	hypothetical prote
467	6	3.1	727	2	G01792	transcription fact	540	6	3.1	969	2	H69425	modification methy

249	3.1	354	2	E69499	iron-sulfur bindin	322	6	3.1	444	2	B96498	probable polygalac
250	3.1	355	1	RGHUI2	GTP-binding regula	323	6	3.1	444	2	C96498	probable polygalac
251	3.1	355	1	RGRTI2	GTP-binding regula	324	6	3.1	444	2	S54039	hypothetical prote
252	3.1	355	1	RGMSI2	GTP-binding regula	325	6	3.1	447	2	S52391	serotonomin B - mo
253	3.1	355	2	A61031	GTP-binding regula	326	6	3.1	449	2	S02011	hypothetical prote
254	3.1	355	2	I50238	Gi2 protein alpha-	327	6	3.1	449	2	JQ1277	triacylglycerol li
255	3.1	355	2	S28158	GTP-binding regula	328	6	3.1	450	2	T16220	hypothetical prote
256	3.1	355	2	T19016	hypothetical prote	329	6	3.1	451	2	T16418	hypothetical prote
257	3.1	358	2	T37805	hypothetical prote	330	6	3.1	452	2	F90452	conserved hypothet
258	3.1	360	2	T43254	GTP-binding protei	331	6	3.1	454	2	F97918	aspartate kinase I
259	3.1	360	2	D69423	flagellar biosynth	332	6	3.1	454	2	G95047	aspartate kinase I
260	3.1	360	2	S39916	cwlI protein - Bac	333	6	3.1	457	1	A28468	chromogranin A pre
261	3.1	361	2	H86763	thiamin biosynthes	334	6	3.1	457	2	A95940	probable integral
262	3.1	362	2	E71637	hypothetical prote	335	6	3.1	457	2	T40770	coiled coil protei
263	3.1	362	2	G97849	hypothetical prote	336	6	3.1	460	2	A96555	unknown protein li
264	3.1	363	2	AH2721	fatty acid/phospho	337	6	3.1	463	2	A64381	NADH oxidase - Met
265	3.1	370	1	BVLVWX	mbpx protein - liv	338	6	3.1	464	2	AD2289	hypothetical prote
266	3.1	371	2	B49515	phosphoprotein 75	339	6	3.1	485	1	JC1318	triacylglycerol li
267	3.1	373	2	B90512	DNA polymerase III	340	6	3.1	466	2	T41076	hypothetical prote
268	3.1	374	2	H83693	carboxypeptidase G	341	6	3.1	467	2	T39070	probable ubiquitin
269	3.1	375	2	T09374	hypothetical prote	342	6	3.1	469	2	B82446	GGDEF family prote
270	3.1	378	2	H96773	hypothetical prote	343	6	3.1	471	2	S02016	U1 snRNP 70K prote
271	3.1	379	1	E71162	hypothetical prote	344	6	3.1	472	2	T29313	hypothetical prote
272	3.1	379	2	A90337	proteinase, probab	345	6	3.1	472	2	T25949	hypothetical prote
273	3.1	381	2	T24256	hypothetical prote	346	6	3.1	475	2	A81735	cysteinyl-tRNA syn
274	3.1	382	2	D75001	hypothetical prote	347	6	3.1	476	2	A43942	lipase - Pseudomon
275	3.1	383	2	H71848	probable na+/h+ an	348	6	3.1	478	2	S57203	hexokinase - yeast
276	3.1	383	2	G64667	NA+/H+ antiporter	349	6	3.1	481	2	A56346	transcription fact
277	3.1	385	2	JC2256	aspartate transami	350	6	3.1	483	2	D87752	protein C18E3.6 li
278	3.1	386	1	A41691	ubiquinol-cytochro	351	6	3.1	483	2	T15180	hypothetical prote
279	3.1	388	2	T26977	hypothetical prote	352	6	3.1	485	1	K1BYHA	hexokinase (EC 2.7
280	3.1	389	2	T34766	iron-sulfur cofact	353	6	3.1	485	2	T17604	hypothetical prote
281	3.1	391	2	S22579	translation initia	354	6	3.1	485	2	D84475	probable Athila re
282	3.1	391	2	T51737	RNA helicase RH2 I	355	6	3.1	487	2	T34887	probable transposa
283	3.1	391	2	T23292	hypothetical prote	356	6	3.1	487	2	B90276	methylmalonate-sem
284	3.1	393	1	A48357	nonstructural prot	357	6	3.1	492	2	T20368	hypothetical prote
285	3.1	395	2	G82298	tyrosyl-tRNA synth	358	6	3.1	493	2	A12314	hypothetical prote
286	3.1	396	2	E97947	cell division prot	359	6	3.1	494	2	A48133	pre-mRNA splicing
287	3.1	399	2	C83563	tyrosyl-tRNA synth	360	6	3.1	496	2	JH0206	hypothetical 57.4K
288	3.1	399	2	F89978	conserved hypothet	361	6	3.1	496	2	I39478	repR protein - Str
289	3.1	399	2	B95080	cell division prot	362	6	3.1	496	2	S06925	replication initia
290	3.1	400	2	A96996	tyrosyl-tRNA synth	363	6	3.1	496	2	B96674	hypothetical prote
291	3.1	402	2	A48770	probable membrane	364	6	3.1	497	2	D72213	conserved hypothet
292	3.1	403	2	B72778	probable processin	365	6	3.1	498	2	D83994	xylose kinase xylB
293	3.1	403	2	T38078	hypothetical prote	366	6	3.1	501	2	S16711	ABC1 protein precu
294	3.1	404	2	A96498	probable polygalac	367	6	3.1	501	2	T21973	hypothetical prote
295	3.1	404	2	H87635	Rieske 2Fe-2S fami	368	6	3.1	502	2	T29729	hypothetical prote
296	3.1	406	2	F90935	acetylornithine de	369	6	3.1	503	2	F82275	toxin co-regulated
297	3.1	406	2	B85784	succinylornithine de	370	6	3.1	503	2	JN0524	hypothetical prote
298	3.1	406	2	D64934	tyrosine-tRNA liga	371	6	3.1	504	2	T31784	pyruvate kinase (E
299	3.1	407	1	A55515	translation initia	372	6	3.1	508	2	H64223	protein F14N23.31
300	3.1	407	2	T06824	MSHA biogenesis pr	373	6	3.1	512	2	A86238	probable structure
301	3.1	407	2	C82327	hypothetical prote	374	6	3.1	512	2	T40576	E2 protein - human
302	3.1	410	2	E72397	hypothetical prote	375	6	3.1	514	1	W2ML5	E2 protein - human
303	3.1	411	2	T25223	3-oxoacyl-l-acyl-ca	376	6	3.1	515	1	F88618	protein W06F12.2 I
304	3.1	416	2	AH2223	MYB DNA-binding-li	377	6	3.1	515	2	J50083	cellulose 1,4-beta
305	3.1	423	2	T51794	transcription init	378	6	3.1	516	2	S33164	cellulose 1,4-beta
306	3.1	424	2	T14358	transcription init	379	6	3.1	516	2	S60661	legumin - Weiwitsc
307	3.1	424	2	T14359	hypothetical prote	380	6	3.1	519	2	S35575	myosin heavy chain
308	3.1	425	2	T43772	multifunctional pu	381	6	3.1	520	2	G84510	probable receptor-
309	3.1	425	2	S14147	SH3-domain binding	382	6	3.1	520	2	T52584	diacylglycerol O-a
310	3.1	425	2	JEO086	aminoimidazole rib	383	6	3.1	520	2	G84864	probable membrane
311	3.1	425	2	S55684	5-aminimidazole r	384	6	3.1	521	2	T05994	protein kinase hom
312	3.1	426	2	A35641	probable chromatin	385	6	3.1	523	2	H84510	cellulose 1,4-beta
313	3.1	431	2	T38471	hypothetical prote	386	6	3.1	524	2	S38794	hypothetical prote
314	3.1	431	2	T33287	hypothetical prote	387	6	3.1	525	1	T49212	hypothetical prote
315	3.1	438	2	T47711	ankyrin-repeat pro	388	6	3.1	527	2	T10388	hypothetical prote
316	3.1	439	2	JQ1729	diacylglycerol O-a	389	6	3.1	530	2	H72864	AcOrf-119 protein
317	3.1	441	2	H84576	hypothetical prote	390	6	3.1	542	2	D87185	probable UDP-gluc
318	3.1	441	2	T34401	hypothetical prote	391	6	3.1	545	2	AI2327	hypothetical prote
319	3.1	443	2	S37915	CBS domain protein	392	6	3.1	548	2	A56353	MIDI protein - yea
320	3.1	443	2	E95229	hemolysin [importe	393	6	3.1	549	2	D87000	acetyl/propionyl C
321	3.1	443	2	H98093		394	6	3.1				

103	6	3.1	168	2	B97776	hypothetical prote	176	6	3.1	292	2	G88448	protein C45G9.4 [i
104	6	3.1	170	2	D86522	CT144 hypothetical	177	6	3.1	294	2	A44856	haloacetate dehalo
105	6	3.1	171	2	T48393	hypothetical prote	178	6	3.1	294	2	H96662	hypothetical prote
106	6	3.1	173	2	G81428	hypothetical prote	179	6	3.1	294	2	E88448	protein C45G9.9 [i
107	6	3.1	174	2	A11979	hypothetical prote	180	6	3.1	296	2	S60251	mab-18 protein (tr
108	6	3.1	180	2	A81287	ribosomal-protein-	181	6	3.1	296	2	AG3515	N-formylglutamate
109	6	3.1	181	2	T48241	hypothetical prote	182	6	3.1	297	2	B58460	phosphoprotein NS
110	6	3.1	187	1	A45113	ribosomal protein	183	6	3.1	298	2	A70238	hypothetical prote
111	6	3.1	189	2	E90157	conserved hypothet	184	6	3.1	299	2	A13184	transcription regu
112	6	3.1	191	2	D64418	hypothetical prote	185	6	3.1	299	2	E84262	rhannosyl transfer
113	6	3.1	196	2	G95114	exonuclease (impor	186	6	3.1	303	1	TVH0UD	transforming prote
114	6	3.1	197	2	T33525	hypothetical prote	187	6	3.1	303	2	T15694	hypothetical prote
115	6	3.1	198	2	F71181	hypothetical prote	188	6	3.1	303	2	T23583	hypothetical prote
116	6	3.1	201	2	T42647	hypothetical prote	189	6	3.1	304	2	T27583	hypothetical prote
117	6	3.1	202	1	R31V4	ribosomal protein	190	6	3.1	305	2	B72691	probable fumarate
118	6	3.1	202	2	A82378	hypothetical prote	191	6	3.1	306	2	F87346	conserved hypothet
119	6	3.1	203	2	B70254	conserved hypothet	192	6	3.1	307	2	T47676	zinc finger protei
120	6	3.1	205	2	S34919	ribosomal protein	193	6	3.1	309	2	S77170	site-specific DNA-
121	6	3.1	206	1	E64317	hypothetical prote	194	6	3.1	310	2	E90231	UDP-glucose 4-epim
122	6	3.1	206	2	T36820	hypothetical prote	195	6	3.1	312	2	F86274	hypothetical prote
123	6	3.1	207	2	T24858	hypothetical prote	196	6	3.1	314	2	T46743	carbamate kinase (
124	6	3.1	208	2	H97983	DNA-directed DNA p	197	6	3.1	314	2	AC1484	B. subtilis Yqfa p
125	6	3.1	210	2	A65009	thiamin phosphate	198	6	3.1	314	2	A11123	B. subtilis Yqfa p
126	6	3.1	214	2	T52495	hypothetical prote	199	6	3.1	315	2	H82512	probable 3-hydroxy
127	6	3.1	221	2	H97021	hypothetical prote	200	6	3.1	315	2	T05512	hypothetical prote
128	6	3.1	222	2	S26994	cellulose 1,4-beta	201	6	3.1	317	2	P00339	multimeric flavodo
129	6	3.1	224	2	C81417	probable transcrip	202	6	3.1	317	2	E97211	hypothetical prote
130	6	3.1	226	2	T29210	hypothetical prote	203	6	3.1	318	2	G96910	D-3 phosphoglycera
131	6	3.1	227	2	C72559	hypothetical prote	204	6	3.1	319	2	A69756	adhesion protein h
132	6	3.1	228	2	T23994	hypothetical prote	205	6	3.1	321	2	G95920	probable epimerase
133	6	3.1	230	2	E86765	hypothetical prote	206	6	3.1	321	2	G84750	probable homeodoma
134	6	3.1	232	2	T21526	hypothetical prote	207	6	3.1	321	2	T16451	hypothetical prote
135	6	3.1	236	2	E70028	conserved hypothet	208	6	3.1	323	2	S20099	transforming prote
136	6	3.1	239	2	F75111	hypothetical prote	209	6	3.1	325	2	AB1123	B. subtilis lolC p
137	6	3.1	242	2	A69026	conserved hypothet	210	6	3.1	325	2	AD1483	B. subtilis lolC p
138	6	3.1	243	2	AF1426	reductases homolog	211	6	3.1	326	2	T38861	hypothetical prote
139	6	3.1	244	2	B71022	hypothetical prote	212	6	3.1	327	2	JC1437	glucan endo-1,3-be
140	6	3.1	248	2	H86285	protein F9L1.13 [i	213	6	3.1	328	2	E72347	flagellar motor sw
141	6	3.1	248	2	A49515	phosphoprotein 41	214	6	3.1	329	2	G98165	hypothetical prote
142	6	3.1	249	2	S07917	RF3 protein - yeas	215	6	3.1	330	2	AH3121	conserved hypothet
143	6	3.1	251	2	T22839	hypothetical prote	216	6	3.1	330	2	S58255	hypothetical prote
144	6	3.1	252	2	F70226	conserved hypothet	217	6	3.1	331	2	T13145	cytosine-specific
145	6	3.1	252	2	H90191	hypothetical prote	218	6	3.1	331	2	C84515	hypothetical prote
146	6	3.1	259	2	F71561	probable rRNA meth	219	6	3.1	332	1	J50274	biotin synthase (E
147	6	3.1	261	2	B86643	carbonyl reductase	220	6	3.1	332	2	T32083	hypothetical prote
148	6	3.1	261	2	S43218	hypothetical prote	221	6	3.1	335	2	B86594	yop translocation
149	6	3.1	261	2	T22120	hypothetical prote	222	6	3.1	335	2	F72030	type III secretion
150	6	3.1	262	2	G81954	cytochrome C1 prec	223	6	3.1	338	2	T47920	hypothetical prote
151	6	3.1	262	2	F97151	ABC transported MD	224	6	3.1	341	1	TWNSJD	transforming prote
152	6	3.1	264	2	T50640	zinc finger protei	225	6	3.1	341	2	JC4051	jun-D protein - ra
153	6	3.1	265	2	A64705	hypothetical prote	226	6	3.1	343	2	T24700	hypothetical prote
154	6	3.1	265	2	A71815	hypothetical prote	227	6	3.1	344	2	A69661	transcription regu
155	6	3.1	265	2	S62363	D-2-chloropropioni	228	6	3.1	344	2	F69272	hypothetical prote
156	6	3.1	265	2	T43123	probable positive	229	6	3.1	345	2	T41473	probable zinc fing
157	6	3.1	266	2	D81011	ubiquinol-cytochro	230	6	3.1	345	2	C75207	DNA primase PAB223
158	6	3.1	266	2	T49052	hypothetical prote	231	6	3.1	345	2	C89882	conserved hypothet
159	6	3.1	267	2	A13458	urease accessory p	232	6	3.1	346	2	G96643	hypothetical prote
160	6	3.1	268	2	G96016	probable gluconate	233	6	3.1	347	2	A43815	transforming prote
161	6	3.1	270	2	D69127	hypothetical prote	234	6	3.1	349	2	AB0783	probable membrane
162	6	3.1	271	1	IKECM	colicin M - Escher	235	6	3.1	350	2	AG1721	uroporphyrinogen I
163	6	3.1	272	2	AC2037	hypothetical prote	236	6	3.1	351	2	AC3090	conserved hypothet
164	6	3.1	273	2	T18809	hypothetical prote	237	6	3.1	351	2	G98196	hypothetical prote
165	6	3.1	276	2	S70815	glycosyl transfera	238	6	3.1	352	2	T06747	hypothetical prote
166	6	3.1	277	2	T27601	hypothetical prote	239	6	3.1	353	2	F75562	ornithine carbamoy
167	6	3.1	277	2	T40033	probable mitochond	240	6	3.1	353	2	AD1351	uroporphyrinogen I
168	6	3.1	279	2	H83835	hypothetical prote	241	6	3.1	353	2	E97503	fatty acid/phospho
169	6	3.1	280	1	F64440	conserved hypothet	242	6	3.1	354	1	RGH011	GTP-binding regula
170	6	3.1	280	2	A81027	lacto-N-neotetraos	243	6	3.1	354	1	RGBO11	GTP-binding regula
171	6	3.1	280	2	T09576	lxb1 transcription	244	6	3.1	354	1	RGXT11	GTP-binding regula
172	6	3.1	283	2	T10228	hypothetical prote	245	6	3.1	354	1	RGXLI1	GTP-binding regula
173	6	3.1	286	2	C72456	hypothetical prote	246	6	3.1	354	2	I50237	GTP-binding regula
174	6	3.1	288	2	D84616	probable DHLH tran	247	6	3.1	354	2	S28157	GTP-binding regula
175	6	3.1	292	2	C69106	glucose-1-phosphat	248	6	3.1	354	2	T05975	probable carboxy-t

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2003, 09:58:43 ; Search time 46 Seconds

(without alignments)
401.257 Million cell updates/sec

Title: US-09-868-352-23

Perfect score: 192

Sequence: 1 MKRVEPTEMLNKNYPGQF.....EEDKEMTSKQHLLEVRKN 192

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : PIR_73:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	12	6.2	102	2 F70023	hypothetical prote
2	12	6.2	170	2 H95088	conserved hypotet
3	12	6.2	176	2 C97956	hypothetical prote
4	11	5.7	91	2 AB1375	B. subtilis YutD p
5	11	5.7	91	2 AH1744	B. subtilis YutD p
6	11	5.7	93	2 A84079	hypothetical prote
7	8	4.2	109	2 T23176	hypothetical prote
8	8	4.2	251	2 B86647	hypothetical prote
9	8	4.2	502	2 S64928	hypothetical prote
10	8	4.2	733	2 A57459	ribosomal protein
11	7	3.6	124	2 S73210	ribosomal protein
12	7	3.6	127	2 F89858	conserved hypotet
13	7	3.6	160	2 E70416	hypothetical prote
14	7	3.6	194	2 E71407	hypothetical prote
15	7	3.6	194	2 T05238	hypothetical prote
16	7	3.6	250	2 T47871	hypothetical prote
17	7	3.6	317	2 JE0175	frezzled protein-1
18	7	3.6	344	2 S59043	spilling factor SR
19	7	3.6	364	2 AC3224	hypothetical prote
20	7	3.6	381	2 I39939	phenylalanine dehy
21	7	3.6	382	1 VHIHPC	nucleocapsid prote
22	7	3.6	382	1 S24282	nucleocapsid prote
23	7	3.6	382	2 S03762	nucleocapsid prote
24	7	3.6	382	2 E36607	nucleocapsid prote
25	7	3.6	382	2 S47428	nucleocapsid prote
26	7	3.6	405	2 F75596	hypothetical prote
27	7	3.6	407	2 B70185	probable hydroxyme
28	7	3.6	480	2 T20382	hypothetical prote
29	7	3.6	484	2 S67625	probable membrane

30	7	3.6	497	2 T29814	hypothetical prote
31	7	3.6	561	2 A87627	PDZ domain family
32	7	3.6	577	2 S64250	probable membrane
33	7	3.6	599	2 T16774	hypothetical prote
34	7	3.6	628	2 B86789	hypothetical prote
35	7	3.6	632	2 F84996	phosphotransferase
36	7	3.6	729	2 A97747	oligopeptidase B (
37	7	3.6	788	1 I59282	diacylglycerol kin
38	7	3.6	791	2 A53691	diacylglycerol kin
39	7	3.6	798	2 JN0878	100K protein - fow
40	7	3.6	819	2 C84615	hypothetical prote
41	7	3.6	833	2 A90575	conserved hypotet
42	7	3.6	840	2 T01357	hypothetical prote
43	7	3.6	859	2 F84981	leucine-tRNA ligas
44	7	3.6	895	2 T32374	hypothetical prote
45	7	3.6	911	2 T01353	probable serine/th
46	7	3.6	1017	2 T30195	probable serine/th
47	7	3.6	1025	2 A83186	DNA-directed DNA p
48	7	3.6	1231	2 S30185	probable RND efflu
49	7	3.6	1240	2 JC5209	insulin receptor s
50	7	3.6	1242	2 JS0670	insulin receptor s
51	7	3.6	1318	1 HIBPD7	internal virion pr
52	7	3.6	2386	2 T39911	rad3 checkpoint pr
53	7	3.6	3006	2 T28625	variant-specific s
54	6	3.1	35	2 B45316	sperm-specific pro
55	6	3.1	56	2 H82647	hypothetical prote
56	6	3.1	60	2 A56547	sex-peptide precur
57	6	3.1	64	2 G70731	hypothetical prote
58	6	3.1	75	2 F69139	ribosomal protein
59	6	3.1	84	2 G90044	hypothetical prote
60	6	3.1	92	2 S34115	sperm-specific pro
61	6	3.1	102	2 F81377	flagellar motor sw
62	6	3.1	103	2 C97755	cyay protein limpo
63	6	3.1	107	2 F72059	hypothetical prote
64	6	3.1	107	2 E86565	hypothetical prote
65	6	3.1	107	2 F86048	hypothetical prote
66	6	3.1	107	2 B91202	hypothetical prote
67	6	3.1	109	2 G82660	hypothetical prote
68	6	3.1	111	2 S23601	hypothetical prote
69	6	3.1	112	2 A27887	Ig kappa chain V r
70	6	3.1	112	2 S53881	citr protein - Lac
71	6	3.1	114	2 B87129	hypothetical prote
72	6	3.1	118	2 G90227	conserved hypotet
73	6	3.1	120	2 S47819	hypothetical 13.9K
74	6	3.1	120	2 B91188	hypothetical prote
75	6	3.1	120	2 B86035	hypothetical prote
76	6	3.1	122	2 T15766	hypothetical prote
77	6	3.1	122	2 AE0573	conserved hypotet
78	6	3.1	123	1 HSTR2B	histone H2B - rain
79	6	3.1	125	1 HSLB1	histone H2B.1 - Af
80	6	3.1	125	1 HSLB2	histone H2B.2 - Af
81	6	3.1	126	2 JH0362	histone H2B.V - ch
82	6	3.1	126	2 S33220	histone H2B.A - Af
83	6	3.1	126	2 S33221	histone H2B.B - Af
84	6	3.1	126	2 IS1446	histone H2B - Afri
85	6	3.1	126	2 T06861	ribosomal protein
86	6	3.1	127	2 AF1253	hypothetical prote
87	6	3.1	131	2 T01293	hypothetical prote
88	6	3.1	132	2 D86885	30S ribosomal prot
89	6	3.1	144	2 H81570	conserved hypotet
90	6	3.1	148	2 F75091	hypothetical prote
91	6	3.1	149	2 G86521	ribosomal protein
92	6	3.1	149	2 C72101	ribosomal protein
93	6	3.1	150	2 A71554	ribosomal protein
94	6	3.1	150	2 G81705	ribosomal protein
95	6	3.1	150	2 T38477	very hypotet
96	6	3.1	155	2 B71223	hypothetical prote
97	6	3.1	155	2 C75191	transcription regu
98	6	3.1	156	2 T32427	hypothetical prote
99	6	3.1	158	2 D81714	conserved hypotet
100	6	3.1	166	2 S59922	allergen RA14B pre
101	6	3.1	168	2 T21692	hypothetical prote
102	6	3.1	168	2 T08831	disease resistance

THIS PAGE BLANK (LEFT)

SQ SEQUENCE 367 AA; 40176 MW; ABC91F55B84536EE CRC64;

Query Match 3.6%; Score 7; DB 10; Length 367;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 FIHFENI 26

|||||

Db 260 FIHFENI 266

RESULT 40

O12298

ID O12298 PRELIMINARY; PRT; 376 AA.

AC O12298;

DT 01-JUL-1997 (TREMBLrel. 04, Created)

DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE Nucleocapsid protein.

GN N.

OS feline coronavirus

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;

OC Coronaviridae; Coronavirus.

OX NCBI_TaxID=12663;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=PECV 79-1683;

RA Herrewegh A.A.P.M., Veldkamp A., de Rooij E., Vennema H.,

RA de Groot R.J.;

RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL: Y13921; CAA74230.1; ..

DR InterPro: IPR001218; Corona.nucleocap.

DR Pfam: PF00937; Corona.nucleoca; 1.

SQ SEQUENCE 376 AA; 42395 MW; F9A8C22350E1D6FD CRC64;

Query Match

Best Local Similarity 3.6%; Score 7; DB 12; Length 376;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 RKSRSKS 135

|||||

Db 338 RKSRSKS 344

Search completed: February 26, 2003, 10:01:10

Job time : 67 secs

```

RESULT 36
Q921K3 ID Q921K3 PRELIMINARY; PRT; 339 AA.
AC Q921K3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2002 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE RIKEN cDNA 1210001E11 gene.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC012039; AAH12039.1; -.
DR InterPro: IPR000504; RNA_rec_mot.
DR Pfam: PF000076; rrm; 2.
DR PROSITE: PS0102; RRM; 2.
SQ SEQUENCE 339 AA; 39029 MW; AE8B6650D0448A0D CRC64;

Query Match 3.6%; Score 7; DB 11; Length 339;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 RKSRKS 135
DB 241 RKSRKS 247
|||||

RESULT 37
Q93ED3 ID Q93ED3 PRELIMINARY; PRT; 358 AA.
AC Q93ED3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2002 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE IMPA.
GN IMPA.
OS Rhizobium leguminosarum (biovar trifolii).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=386;
RN [1]
RP SEQUENCE FROM N.A.
RA Bladergroen M.R.; Badelt K.; Stronk O.P.; Lugtenberg E.J.J.;
RA Spaijk H.P.;
RT "An avirulence locus of a symbiotic Rhizobium leguminosarum strain is
RT involved in temperature-dependent protein secretion."
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF361470; AAL17808.1; -.
SQ SEQUENCE 358 AA; 39039 MW; 9F5C107C58D22395 CRC64;

Query Match 3.6%; Score 7; DB 2; Length 358;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 QLRLRF 77
DB 87 QLRLRF 93
|||||

RESULT 38
Q8UJF4 ID Q8UJF4 PRELIMINARY; PRT; 364 AA.
AC Q8UJF4;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE ABC transporter, membrane spanning protein.
GN ATU5523 OR AGR_PAT_778.

```

```

OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OG Plasmid AT.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendenning J., Decherage G., Gillet W., Grant C.,
RA Kutayin T., Levy R., Li M.-J., McClelland E., Palmeri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58."
RL Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RA MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmlel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gursion J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58."
RL Science 294:2323-2328(2001).
DR EMBL: AE008972; AAL46209.1; -.
DR EMBL: AE007920; AAK90899.1; -.
KW Plasmid; Complete proteome.
SQ SEQUENCE 364 AA; 38498 MW; F719F8667AA3CE93 CRC64;

Query Match 3.6%; Score 7; DB 16; Length 364;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 LRLRGFY 78
DB 117 LRLRGFY 123
|||||

RESULT 39
Q92PI8 ID Q92PI8 PRELIMINARY; PRT; 367 AA.
AC Q92PI8;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Polygalacturonase.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RA Bergey D.R., Orozco-Cardenas M., de Moura D.S., Ryan C.A.;
RT "A wound- and systemin-inducible polygalacturonase in tomato leaves."
RL Proc. Natl. Acad. Sci. U.S.A. 96:1756-1760(1999).
CC -!- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES
(C POLYGLACTURONASES).
DR EMBL: AF118567; AAD17250.1; -.
DR InterPro: IPR000743; GH28.
DR Pfam: PF00295; Glyco_hydro_28; 1.
DR PROSITE: PS00502; POLYGLACTURONASE; 1.
KW Cell wall; Glycosidase; Hydrolase.

```

```

RA Meikonyan H., Chang W.C., Shapiro J.P., Mahadevappa M.,
RA Fitzpatrick P.A., Kiefer M.C., Tomei D.L., Umansky S.R.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Chang J.T., Esumi N., Moore K., Li Y., Zhang S., Chew C., Goodman B.,
RA Rattner A., Moody S., Stetten G., Campochiaro P.A., Zack D.J.;
RT "Cloning and characterization of a secreted frizzled-related protein
RT that is expressed by the retinal pigment epithelium.";
RL Hum. Mol. Genet. 0:0-0(1999).
DR EMBL; AF017988; AAB/0794.1; -.
DR EMBL; AF117758; AAD/25052.1; -.
DR InterPro; IPR000024; Fz_domain.
DR InterPro; IPR001134; Netrin_C.
DR Pfam; PF01392; Fz; 1.
DR Pfam; PF01759; NTR; 1.
DR SMART; SM00063; FRI; 1.
DR PROSITE; PS00038; FZ; 1.
SQ SEQUENCE 317 AA; 35577 MW; 05F1FCCCB614F36A CRC64;

Query Match 3.6%; Score 7; DB 4; Length 317;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 159 KRKDKTKR 165
Db 236 KRKDKTKR 242
|||||

RESULT 33
Q96I27 PRELIMINARY; PRT; 334 AA.
AC Q96I27;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Similar to RIKEN cDNA 1200013F24 gene.
OS Homo sapiens (Human).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BLADDER;
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC006982; AAH06982.1; -.
DR InterPro; IPR001993; Mitoch_carrier.
DR PROSITE; PS00215; MITOCH_CARRIER; UNKNOWNL1.
SQ SEQUENCE 334 AA; 38677 MW; DF31CC8C953E790A CRC64;

Query Match 3.6%; Score 7; DB 4; Length 334;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 129 RKSRKS 135
Db 42 RKSRKS 48
|||||

RESULT 34
Q96FN8 PRELIMINARY; PRT; 337 AA.
AC Q96FN8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Unknown (protein for MGC:16799).
OS Homo sapiens (Human).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC TISSUE=EYE;
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC010619; AAHL0619.1; -.
DR InterPro; IPR005302; MOSC.
DR InterPro; IPR005303; MOSC_N.
DR Pfam; PF03473; MOSC; 1.
DR Pfam; PF03476; MOSC_N; 1.
SQ SEQUENCE 337 AA; 37466 MW; 635967B8FE1C8374 CRC64;

Query Match 3.6%; Score 7; DB 4; Length 337;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 107 YFVLENP 113
Db 317 YFVLENP 323
|||||

RESULT 35
Q9DBP1 PRELIMINARY; PRT; 339 AA.
AC Q9DBP1;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE 1210001E1IRIK protein.
GN 1210001E1IRIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LUNG;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaudo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK004831; BAB23599.1; -.
DR MGD; MGI:1915246; 1210001E1IRik.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 2.
DR SMART; SM00360; RRM; 2.
DR PROSITE; PS0102; RRM; 2.
SQ SEQUENCE 339 AA; 38997 MW; BC3C9BDA38144A0D CRC64;

Query Match 3.6%; Score 7; DB 11; Length 339;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 129 RKSRKS 135
Db 241 RKSRKS 247
|||||

```


DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE 1210001E11Rik protein.
 GN 1210001E11Rik.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10990;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=EMBRYO;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
 RA Hayashizaki Y.,
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 DR EMBL; AK012605; BAB28351.1; -;
 DR MGD; MGI:1915246; 1210001E11Rik.
 DR InterPro; IPR000504; RNA_rec_mot.
 DR Pfam; PF000076; rtm; 1.
 DR SMART; SM00360; RRM; 1.
 DR PROSITE; PS50102; RRM; 1.
 SQ SEQUENCE 305 AA; 34515 MW; 95F8D2E3D921E500 CRC64;

 Query Match 3.6%; Score 7; DB 11; Length 305;
 Best Local Similarity 100.0%; Pred. No. 58;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 129 RKSRSKS 135
 |||||
 DB 207 RKSRSKS 213

 RESULT 30
 Q9PVZ0 PRELIMINARY; PRT; 314 AA.
 ID Q9PVZ0
 AC Q9PVZ0;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
 DE C-Jun protein.
 GN C-JUN.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20500889; PubMed=11044605;
 RA Knoechel S., Schuler-Wetz A., Knoechel W.;
 RT "c-Jun (AP-1) activates BMP-4 transcription in Xenopus embryos.";
 RL Mech. Dev. 98:29-36(2000).
 CC -|- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 CC -|- SIMILARITY: BELONGS TO THE BZIP FAMILY.
 DR EMBL; AJ243955; CAB51637.1; -;
 DR HSP; P05412; 1F0S.

DR InterPro; IPR002112; Leuzip_Jun.
 DR InterPro; IPR004827; TF_bZIP.
 DR Pfam; PF00170; bZIP; 1.
 DR PRINTS; PR00043; LEUZIPPRJUN.
 DR SMART; SM00338; BRL2; 1.
 DR PROSITE; PS00036; BZIP_BASIC; 1.
 KW DNA-binding; Nuclear protein.
 SQ SEQUENCE 314 AA; 34556 MW; 73396B48ACAC93B3 CRC64;

 Query Match 3.6%; Score 7; DB 13; Length 314;
 Best Local Similarity 100.0%; Pred. No. 60;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 89 RISRLD 95
 |||||
 DB 259 RISRLD 265

 RESULT 31
 Q9XSC1 PRELIMINARY; PRT; 315 AA.
 ID Q9XSC1
 AC Q9XSC1;
 DT 01-NOV-1999 (Tremblrel. 12, Created)
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Secreted frizzled-related protein 5.
 GN SFRP5.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Chang J.T., Esumi N., Moore K., Li Y., Zhang S., Chew C., Goodman B.,
 RA Rattner A., Moody S., Stetten G., Campochiaro P.A., Zack D.J.;
 RT "Cloning and characterization of a secreted frizzled-related protein
 RT that is expressed by the retinal pigment epithelium."
 RL Hum. Mol. Genet. 0:0-0(1999).
 DR EMBL; AF117757; AAD25051.1; -;
 DR InterPro; IPR000024; Fz_domain.
 DR InterPro; IPR001134; Netrin_C.
 DR Pfam; PF01392; Fz; 1.
 DR Pfam; PF01759; NTR; 1.
 DR SMART; SM00063; FRI; 1.
 DR PROSITE; PS00038; FZ; 1.
 SQ SEQUENCE 315 AA; 35279 MW; 8B6E94951060A976 CRC64;

 Query Match 3.6%; Score 7; DB 6; Length 315;
 Best Local Similarity 100.0%; Pred. No. 60;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 159 KRKDKR 165
 |||||
 DB 234 KRKDKR 240

 RESULT 32
 O14780 PRELIMINARY; PRT; 317 AA.
 ID O14780
 AC O14780;
 DT 01-JAN-1998 (Tremblrel. 05, Created)
 DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Secreted apoptosis related protein 3.
 GN SARP3 OR SFRP5.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PANCREAS;

```

RESULT 25
Q8V295 PRELIMINARY; PRT; 256 AA.
AC Q8V295;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical 28.5 kDa protein.
GN AT3G60600, T4C21_10.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY065144; AAL38320.1; -.
DR InterPro; IPR000535; MSP_domain.
DR Pfam; PF00635; MSP_domain; 1.
KW Hypothetical protein.
SQ SEQUENCE 256 AA; 28472 MW; 2D17D79F0F736068 CRC64;

Query Match 3.6%; Score 7; DB 10; Length 256;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 KEVTPM 9
Db 117 KEVTPM 123

RESULT 26
Q960K2 PRELIMINARY; PRT; 276 AA.
AC Q960K2;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Similar to hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RC TISSUE=PROSTATE;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC010357; AAH0357.1; -.
DR InterPro; IPR001993; Mitoch_carrier.
DR PROSITE; PS00215; MITOCH_CARRIER; UNKNOWN_1.
SQ SEQUENCE 276 AA; 31528 MW; 99834B6C3BAF02A CRC64;

Query Match 3.6%; Score 7; DB 4; Length 276;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 RKSRSKS 135
Db 42 RKSRSKS 48

RESULT 27
Q9BI97 PRELIMINARY; PRT; 282 AA.
ID Q9BI97

```

```

AC Q9BI97;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE F17C11.11 protein.
GN F17C11.11.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA McMurray A.A.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL; Z72507; CAC35886.1; -.
SQ SEQUENCE 282 AA; 33577 MW; 212549C6C67EC35E CRC64;

Query Match 3.6%; Score 7; DB 5; Length 282;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 VFGQRF 54
Db 148 VFGQRF 154

RESULT 28
Q8R8H2 PRELIMINARY; PRT; 283 AA.
ID Q8R8H2
AC Q8R8H2;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Antirestriction protein.
GN TTE2027.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WB4T / JCM11007;
RX MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002).
DR EMBL; AE013153; AAM25204.1; -.
KW Complete proteome.
SQ SEQUENCE 283 AA; 32618 MW; 62611FC401EDB312 CRC64;

Query Match 3.6%; Score 7; DB 16; Length 283;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 39 NEKSAFD 45
Db 203 NEKSAFD 209

RESULT 29
Q9CZH6 PRELIMINARY; PRT; 305 AA.
ID Q9CZH6
AC Q9CZH6;
DT 01-JUN-2001 (TReMBLrel. 17, Created)

```

RN [3]
 RP SEQUENCE FROM N.A.
 RA Weber N., Grueninger D., Schmidheini T., Mewes H.W., Lemcke K.,
 RA Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL035528; CAB36833.1; -;
 DR EMBL; AL161537; CAB78411.1; -;
 KW Hypothetical protein.
 SQ SEQUENCE 194 AA; 22117 MW; A50E62552E730F99 CRC64;

 Query Match 3.6%; Score 7; DB 10; Length 194;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 156 TSKKRD 162
 Db 98 TSKKRD 104

 RESULT 22
 Q8SW53
 ID Q8SW53 PRELIMINARY; PRT; 200 AA.
 AC Q8SW53;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Hypothetical protein ECU03_0570.
 GN ECU03_0570.
 OS Encephalitozoon cuniculi.
 OC Eukaryota; Microsporidia; Unikaryonidae; Encephalitozoon.
 OX NCBI_TaxID=6035;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GB-M1;
 RA Genoscope;
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GB-M1;
 RX MEDLINE=21576510; PubMed=11719806;
 RA Katinka M.D., Duprat S., Cornillot E., Metenier G., Thomarat F.,
 RA Prensier G., Barbe V., Peyretailade E., Brottier P., Wincker P.,
 RA Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,
 RA Weissenbach J., Vivares C.P.;
 RT "Genome sequence and gene compaction of the eukaryote parasite
 RT Encephalitozoon cuniculi."
 RL Nature 414:450-453(2001).
 DR EMBL; AL590443; CAD26203.1; -;
 KW Hypothetical protein.
 SQ SEQUENCE 200 AA; 23510 MW; 9DB2F9A4ED2E032C CRC64;

 Query Match 3.6%; Score 7; DB 5; Length 200;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 129 RKSRKS 135
 Db 191 RKSRKS 197

 RESULT 23
 Q9M003
 ID Q9M003 PRELIMINARY; PRT; 250 AA.
 AC Q9M003;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
 DE Hypothetical 27.7 kDa protein.
 GN T4C21_10.

OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Choisne N., Robert C., Brottier P., Wincker P., Cattolico L.,
 RA Artiguenave F., Saurin W., Weissenbach J., Mewes H.W., Rudd S.,
 RA Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL162295; CAB82664.1; -;
 DR InterPro; IPR000535; MSP_domain.
 DR Pfam; PF00635; MSP_domain; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 250 AA; 27705 MW; E8F8DC2F4D685B66 CRC64;

 Query Match 3.6%; Score 7; DB 10; Length 250;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 3 KEVTPEM 9
 Db 117 KEVTPEM 123

 RESULT 24
 Q9NL00
 ID Q9NL00 PRELIMINARY; PRT; 253 AA.
 AC Q9NL00;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE HgTPR5b protein (Fragment).
 GN HGTPR5B.
 OS Eptatretus burgeri (Inshore hagfish).
 OC Eukaryota; Metazoa; Chordata; Craniata; Hyperotreti; Myxiniiformes;
 OC Myxiniidae; Eptatretinae; Eptatretus.
 OX NCBI_TaxID=7764;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20219325; PubMed=10754074;
 RA Ono-Koyanagi K., Suga H., Katoh K., Miyata T.;
 RT "Protein tyrosine phosphatases from amphioxus, hagfish, and ray:
 RT divergence of tissue-specific isoform genes in the early evolution of
 RT vertebrates."
 RL J. Mol. Evol. 50:302-311(2000).
 DR EMBL; AB033577; BAA95184.1; -;
 DR HSP; P28827; 1RPM.
 DR InterPro; IPR000387; TYR_phosphatase.
 DR InterPro; IPR000242; Tyr_PP.
 DR Pfam; PF00102; Y_phosphatase; 1.
 DR PRINTS; PR00700; PRTYPHPTASE.
 DR SMART; SM00194; PTPC; 1.
 DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
 DR PROSITE; PS00566; TYR_PHOSPHATASE_2; 1.
 DR PROSITE; PS00555; TYR_PHOSPHATASE_PTP; 1.
 KW Hydrolase.
 FT NON_TER
 SQ SEQUENCE 253 AA; 28315 MW; 27845DD6FC154C96 CRC64;

 Query Match 3.6%; Score 7; DB 13; Length 253;
 Best Local Similarity 100.0%; Pred. No. 50;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 137 SSKSQTR 143
 Db 122 SSKSQTR 128

RA Graham D.E., Overbeek R., Snead M.A., Keller M., AuJay M., Huber R.,
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
 RT "The complete genome of the hyperthermophilic bacterium Aquifex
 RT aeolicus";
 RL Nature 392:353-358(1998).
 DR EMBL: AE000735; AAC07319.1; -;
 DR InterPro: IPR000063; ThioRed.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 160 AA; 18519 MW; BCC53BFD9D002201 CRC64;

Query Match 3.6%; Score 7; DB 16; Length 160;
 Best Local Similarity 100.0%; Pred. No. 33;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 92 RLEDYIK 98
 |||||
 Db 151 RLEDYIK 157

RESULT 18
 Q9NZE5 PRELIMINARY; PRT; 173 AA.
 ID Q9NZE5
 AC Q9NZE5;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
 DE BM-011.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BONE MARROW;
 RA Zhao M., Song H., Li N., Peng Y., Han Z., Chen Z.;
 RT "A novel gene expressed in human bone marrow";
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF208853; AAF64267.1; -;
 SQ SEQUENCE 173 AA; 20900 MW; 20E3D2AE177DBCC3 CRC64;

Query Match 3.6%; Score 7; DB 4; Length 173;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 129 RKSRKS 135
 |||||
 Db 42 RKSRKS 48

RESULT 19
 Q939N6 PRELIMINARY; PRT; 176 AA.
 ID Q939N6
 AC Q939N6;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Hypothetical 19.0 kDa protein.
 OS Streptococcus gordonii.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
 OC Streptococcaceae; Streptococcus.
 OX NCBI_TaxID=1302;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-M99;
 RA Bensing B.A., Sullam P.M.;
 RT "An accessory sec locus of Streptococcus gordonii is required for
 RT export of GspB and for platelet binding";
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY028381; AAL13052.1; -;
 KW Hypothetical protein.
 SQ SEQUENCE 176 AA; 19006 MW; 16FC686B3B98297D CRC64;

Query Match 3.6%; Score 7; DB 2; Length 176;

Best Local Similarity 100.0%; Pred. No. 36;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 17 GPQFIHF 23
 |||||
 Db 33 GPQFIHF 39

RESULT 20
 O23308 PRELIMINARY; PRT; 194 AA.
 ID O23308
 AC O23308;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE Hypothetical 22.0 kDa protein.
 GN AT4G14520.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bevan M., Stiekema W., Murphy G., Wambutt R., Pohl T., Terry N.,
 RA Kreis M., Kavanagh T., Entian K.D., Rieger M., James R.,
 RA Puigdomenech P., Katzopoulos P., Obermaier B., Duesterhoft A.,
 RA Jones J., Palme K., Ansorge W., Deiseny M., Bancroft I., Mewes H.W.,
 RA Schueller C., Chaiwatzis N.;
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: Z97336; CAB10231.1; -;
 DR EMBL: AL161539; CAB78494.1; -;
 DR InterPro: IPR003029; S1.
 DR SMART: SM00316; S1; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 194 AA; 22026 MW; 89822853E434AF10 CRC64;

Query Match 3.6%; Score 7; DB 10; Length 194;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 123 ERPHKR 129
 |||||
 Db 160 ERPHKR 166

RESULT 21
 Q9SVP3 PRELIMINARY; PRT; 194 AA.
 ID Q9SVP3
 AC Q9SVP3;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE Hypothetical 22.1 kDa protein.
 GN F18A5.80 OR AT4G13690.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bevan M., Weber N., Grueninger D., Schmidheini T., Bancroft I.,
 RA Mewes H.W., Mayer K.F.X., Lemcke K., Schueller C.;
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.

```
DR EMBL; AF003131; BAB42025.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 127 AA; 15401 MW; 4901EFB0676AE64E CRC64;

Query Match          3.6%; Score 7; DB 16; Length 127;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 EYCNFGC 105
Db 75 EYCNFGC 81

RESULT 14
Q9QF07 PRELIMINARY; PRT; 129 AA.
AC Q9QF07;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE GAG polyprotein (fragment).
GN GAG.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LN24;
RX MEDLINE=99412391; PubMed=10482626;
RA Morris A., Marsden M., Halcrow K., Hughes E.S., Brettie R.P.,
RA Bell J.E., Simmonds P.;
RT "Mosaic structure of the human immunodeficiency virus type 1 genome
RT infecting lymphoid cells and the brain: evidence for frequent in vivo
RT recombination events in the evolution of regional populations.";
RL J. Virol. 73:8720-8731(1999).
DR EMBL; AF174941; AAF00287.1; -.
DR InterPro; IPR000071; Retrovir_p17.
DR Pfam; PF00540; Gag_p17; 1.
DR PRINTS; PR00234; HIVMATRIX.
KW AIDS; Core protein; Polyprotein.
FT NON_TER 1
FT NON_TER 129
SQ SEQUENCE 129 AA; 14448 MW; C2234C7181D896B3 CRC64;

Query Match          3.6%; Score 7; DB 15; Length 129;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 IKEEQDK 177
Db 77 IKEEQDK 83

RESULT 15
Q8STE7 PRELIMINARY; PRT; 137 AA.
AC Q8STE7;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE D2089.1b protein.
GN D2089.1b.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Swinburne J.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RX MEDLINE=99069613; PubMed=9851916;
```

```
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
RN [3]
RP SEQUENCE FROM N.A.
RA Gardner A.E.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z36948; CAD30436.1; -.
DR EMBL; Z49911; CAD30436.1; JOINED.
DR EMBL; Z49911; CAD30448.1; -.
DR EMBL; Z36948; CAD30448.1; JOINED.
SQ SEQUENCE 137 AA; 15287 MW; 50F63764F259C1C7 CRC64;

Query Match          3.6%; Score 7; DB 5; Length 137;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 IKFDDER 124
Db 59 IKFDDER 65

RESULT 16
Q9H6C7 PRELIMINARY; PRT; 151 AA.
AC Q9H6C7;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CDNA: FLJ22390 fis, clone HRC07810.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA Suzuki Y., Obayashi M., Nishi T., Shibahara T., Tanaka T.,
RA Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK026043; BAB15333.1; -.
DR InterPro; IPR005302; MOSC.
DR Pfam; PF03473; MOSC; 1.
SQ SEQUENCE 151 AA; 17097 MW; 47B7C93F0ACE74C1 CRC64;

Query Match          3.6%; Score 7; DB 4; Length 151;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 YFVLENP 113
Db 131 YFVLENP 137

RESULT 17
O67360 PRELIMINARY; PRT; 160 AA.
AC O67360;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein AQ_1343.
GN AQ_1343.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
```

Q93U42;
 01-DEC-2001 (TrEMBLrel. 19, Created)
 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 RNA polymerase beta subunit (fragment).
 RPOB.
 OS Carsonella ruddii.
 OC Bacteria; Proteobacteria; gamma subdivision; Candidatus Carsonella.
 OX NCBI_TaxID=114186;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Thao M.L., Clark M.A., Burckhardt D.H., Moran N.A., Baumann P.;
 RT "Phylogenetic analysis of vertically transmitted psyllid endosymbionts
 (Candidatus Carsonella ruddii) based on atpAGD and rpoC; comparisons
 with 16S-23S rDNA-derived phylogeny.";
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
 OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
 SUBSTRATES (BY SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N DIPHOSPHATE +
 [RNA](N).
 CC -!- SUBUNIT: THE ENZYME CONSISTS OF THE SIGMA CHAIN AND THE CORE
 ENZYME WHICH IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, AND 1
 BETA' CHAIN (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.
 DR EMBL: AF268062; AAK55952.1; -;
 DR InterPro: IPR001572; RNA_pol_B.
 DR Pfam: PF00562; RNA_pol_B; 1.
 DR DNA-directed RNA polymerase; Transcription; Transferase.
 FT NON_TER 1
 SQ SEQUENCE 60 AA; 6786 MW; 90BB0074478AFBFD CRC64;
 Query Match 3.6%; Score 7; DB 2; Length 60;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 27 VKSDDIE 33
 Db 13 VKSDDIE 19
 RESULT 11
 ID Q93U40; PRELIMINARY; PRT; 60 AA.
 AC Q93U40;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE RNA polymerase beta subunit (fragment).
 GN RPOB.
 OS Carsonella ruddii.
 OC Bacteria; Proteobacteria; gamma subdivision; Candidatus Carsonella.
 OX NCBI_TaxID=114186;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Thao M.L., Clark M.A., Burckhardt D.H., Moran N.A., Baumann P.;
 RT "Phylogenetic analysis of vertically transmitted psyllid endosymbionts
 (Candidatus Carsonella ruddii) based on atpAGD and rpoC; comparisons
 with 16S-23S rDNA-derived phylogeny.";
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
 OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
 SUBSTRATES (BY SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N DIPHOSPHATE +
 [RNA](N).
 CC -!- SUBUNIT: THE ENZYME CONSISTS OF THE SIGMA CHAIN AND THE CORE
 ENZYME WHICH IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, AND 1
 BETA' CHAIN (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.
 DR EMBL: AF268062; AAK55954.1; -;
 DR InterPro: IPR001572; RNA_pol_B.
 DR Pfam: PF00562; RNA_pol_B; 1.
 DR DNA-directed RNA polymerase; Transcription; Transferase.

FT NON_TER 1
 SQ SEQUENCE 60 AA; 6617 MW; 9A34D68E7E563350 CRC64;
 Query Match 3.6%; Score 7; DB 2; Length 60;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 27 VKSDDIE 33
 Db 13 VKSDDIE 19
 RESULT 12
 ID Q9YNZ7 PRELIMINARY; PRT; 88 AA.
 AC Q9YNZ7;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
 DE Hypothetical 9.6 kDa protein.
 OS Choristoneura fumiferana nuclear polyhedrosis virus (CfMNPV).
 OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
 OC Nucleopolyhedrovirus.
 OX NCBI_TaxID=10448;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=IRELAND;
 RA Poloumienko A., Krell P.J.;
 RT "Identification of the GTA (probable global transactivator) gene of a
 baculovirus pathogenic to the spruce budworm, CfMNPV.";
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U72240; RAD10319.1; -;
 DR Hypothetical protein.
 KW Hypothetical protein.
 SQ SEQUENCE 88 AA; 9631 MW; 0D2042034BCC75F5 CRC64;
 Query Match 3.6%; Score 7; DB 12; Length 88;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 127 KRRKRS 133
 Db 18 KRRKRS 24
 RESULT 13
 ID Q99VF1 PRELIMINARY; PRT; 127 AA.
 AC Q99VF1;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE Hypothetical protein SAV0925.
 GN SAV0925 OR SA0786.
 OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
 OS Staphylococcus aureus (strain N315).
 OC Bacteria; Firmicutes; Bacillus/clostridium group; Bacillales;
 OC Staphylococcus.
 OX NCBI_TaxID=158878, 158879;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S. aureus (strain Mu50), and S. aureus (strain N315);
 RX MEDLINE=2131192; PubMed=11418146;
 RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
 RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
 RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
 RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
 RA Sekinizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
 RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
 RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
 RT "Whole genome sequencing of methicillin-resistant Staphylococcus
 aureus.";
 RL Lancet 357:1225-1240(2001).
 DR EMBL: AP003360; BAB57087.1; -.

DR EMBL; 275543; CAA99865.1; -
SQ SEQUENCE 109 AA; 11870 MW; CE1185AC0F078A47 CRC64;

Query Match 4.2%; Score 8; DB 5; Length 109;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 157 SKKRKDKT 164
Db 73 SKKRKDKT 80

RESULT 7

Q9CJ28 ID Q9CJ28 PRELIMINARY; PRT; 251 AA.
AC Q9CJ28;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Hypothetical protein ybic.
GN YBIC OR LL0178.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Wincker P., Mauer S., Jaillon O., Malarne K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
lactis ssp. lactis IL1403."
RL Genome Res. 11:731-753(2001).
DR EMBL; AE006255; AAK04276.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 251 AA; 30095 MW; 975D661F0C97E01D CRC64;

Query Match 4.2%; Score 8; DB 16; Length 251;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LNYNKYPG 17
Db 10 LNYNKYPG 17

RESULT 8

Q12418 ID Q12418 PRELIMINARY; PRT; 502 AA.
AC Q12418;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Chromosome XII reading frame ORF YLR094C.
GN GIS3 OR L8004.5 OR YLR094C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Benes V., Rechmann S., Nentwich U., Schwager C., Ansoerge W., Voss H.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.

QY 10 LNYNKYPG 17
Db 10 LNYNKYPG 17

RESULT 9

Q96J38 ID Q96J38 PRELIMINARY; PRT; 733 AA.
AC Q96J38;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Ribosomal protein S6 kinase, 90kD, polypeptide 2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=MUSCLE;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.

QY 10 LNYNKYPG 17
Db 10 LNYNKYPG 17

RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA Favello A., Fulton L., Gattung S., Greco T., Kirsten J., Kucaba T.,
RA Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,
RA Johnston L., Langston Y., Latreille P., Le T., Mardis E., Menezes S.,
RA Miller N., Nhan M., Pauley A., Peluso D., Rifken L., Riles L.,
RA Taich A., Trevasakis E., Vignati D., Wilcox L., Wohlman P., Vaudin M.,
RA Wilson R., Waterston R.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.

[5]

RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RA Waterston R.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; 273266; CAA97656.1; -
DR EMBL; 053876; AAB67538.1; -
DR EMBL; 273267; CAA97658.1; -
DR SGD; S0004084; GIS3.
SQ SEQUENCE 502 AA; 56356 MW; 62FADCC7FE59337 CRC64;

Query Match 4.2%; Score 8; DB 3; Length 502;
Best Local Similarity 100.0%; Pred. No. 7.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 RRKRSRKS 135
Db 281 RRKRSRKS 288

RESULT 9

Q96J38 ID Q96J38 PRELIMINARY; PRT; 733 AA.
AC Q96J38;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Ribosomal protein S6 kinase, 90kD, polypeptide 2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=MUSCLE;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.

QY 128 RRKRSRKS 135
Db 19 RRKRSRKS 26

RESULT 10

Q93U42 ID Q93U42 PRELIMINARY; PRT; 60 AA.
QY 128 RRKRSRKS 135
Db 19 RRKRSRKS 26

```
RESULT 3
Q99YU4
ID Q99YU4 PRELIMINARY; PRT; 214 AA.
AC Q99YU4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Hypothetical protein SPY1534.
GN SPY1534.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC 700294 / SEROTYPE M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR EMBL; AE006586; AAK34328.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 214 AA; 25536 MW; B054E21499807357 CRC64;

Query Match 6.2%; Score 12; DB 16; Length 214;
Best Local Similarity 100.0%; Pred. No. 0.00019;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 63 IVGDWGNQQLRL 74
Db 71 IVGDWGNQQLRL 82

RESULT 4
Q928N1
ID Q928N1 PRELIMINARY; PRT; 91 AA.
AC Q928N1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein lmo2402.
GN LMO2402 OR LIN2501.
OS Listeria monocytogenes, and
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1639, 1642;
RN [1]
SEQUENCE FROM N.A.
RC SPECIES=L.monocytogenes, and L.innocua;
RC STRAIN=EGD-E / SEROVAR 1/2A, AND CLIP 11262 / SEROVAR 6A;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkat G.,
RA Madsen E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL591983; CAD00480.1; -.
DR EMBL; AL596172; CAC97728.1; -.
DR ListList; LIN02501; -.

DR ListList; LMO02402; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 91 AA; 11001 MW; AEE305DCE28371C CRC64;

Query Match 5.7%; Score 11; DB 16; Length 91;
Best Local Similarity 100.0%; Pred. No. 0.0011;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 EYCNFGCAYFV 109
Db 76 EYCNFGCAYFV 86

RESULT 5
Q9K7D1
ID Q9K7D1 PRELIMINARY; PRT; 93 AA.
AC Q9K7D1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein BH3433.
GN BH3433.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL; AP001518; BAB07152.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 93 AA; 11230 MW; IAD87F55CA69E6E1 CRC64;

Query Match 5.7%; Score 11; DB 16; Length 93;
Best Local Similarity 100.0%; Pred. No. 0.0011;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 EYCNFGCAYFV 109
Db 75 EYCNFGCAYFV 85

RESULT 6
Q21089
ID Q21089 PRELIMINARY; PRT; 109 AA.
AC Q21089;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE K01D12.9 protein.
GN K01D12.9.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
SEQUENCE FROM N.A.
RC none;
RP Dobson R.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RT none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
```



```

966 6 3.1 982 5 Q960A7
967 6 3.1 982 5 Q9VC93
968 6 3.1 983 5 Q9BI20
969 6 3.1 983 16 Q9HV74
970 6 3.1 985 16 Q92P88
971 6 3.1 1001 5 Q17098
972 6 3.1 1002 3 Q13849
973 6 3.1 1002 10 Q80875
974 6 3.1 1002 10 Q94BS5
975 6 3.1 1004 4 Q9H232
976 6 3.1 1011 5 Q9XVR0
977 6 3.1 1013 3 Q9HF64
978 6 3.1 1020 5 Q9V367
979 6 3.1 1032 16 Q92929
980 6 3.1 1033 5 Q24327
981 6 3.1 1033 5 Q9V643
982 6 3.1 1034 16 Q98IH3
983 6 3.1 1035 5 Q9VJU8
984 6 3.1 1036 16 Q8Y5P1
985 6 3.1 1043 5 Q45232
986 6 3.1 1045 10 Q8RZ01
987 6 3.1 1048 10 Q8W517
988 6 3.1 1050 10 Q81356
989 6 3.1 1052 17 Q8TPC4
990 6 3.1 1053 10 Q9FXK8
991 6 3.1 1054 10 Q9SNY7
992 6 3.1 1057 11 Q91X04
993 6 3.1 1059 3 Q9P8G2
994 6 3.1 1071 16 Q9A0H4
995 6 3.1 1072 11 Q9DBT7
996 6 3.1 1073 16 Q9A4M6
997 6 3.1 1079 10 Q8SLJ6
998 6 3.1 1095 10 Q94J79
999 6 3.1 1095 10 Q943V7
1000 6 3.1 1108 10 Q94ED6

```

ALIGNMENTS

```

RESULT 1
032127 PRELIMINARY; PRT; 102 AA.
ID Q32127
AC Q32127;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DE 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE YUTD protein.
GN YUTD.
OS Bacillus subtilis.
OC Bacteria; Firmicutes.
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrazi E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Gim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guiseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Halbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,

```

```

RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z99120; CAB5221.1; -.
KW Complete proteome.
SQ SEQUENCE 102 AA; 12155 MW; 0551F0BF8AEF3275 CRC64;

Query Match 6.2%; Score 12; DB 16; Length 102;
Best Local Similarity 100.0%; Pred. No. 0.0001;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 EYCNFGCAFEVL 110
DB 86 EYCNFGCAFEVL 97
|||||
QY 99 EYCNFGCAFEVL 110
DB 86 EYCNFGCAFEVL 97

RESULT 2
Q97RN6 PRELIMINARY; PRT; 170 AA.
ID Q97RN6
AC Q97RN6;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Hypothetical protein SP0767.
GN SP0767.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapfel E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae."
RL Science 293:498-506(2001).
DR EMBL; AE007384; AAK74905.1; -.
DR TIGR; SP0767; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 170 AA; 20318 MW; 80273AB3AD61705E CRC64;

Query Match 6.2%; Score 12; DB 16; Length 170;
Best Local Similarity 100.0%; Pred. No. 0.00016;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 NEQLRLRGFYKD 80
DB 69 NEQLRLRGFYKD 80
|||||
QY 69 NEQLRLRGFYKD 80
DB 69 NEQLRLRGFYKD 80

```

820	6	3.1	680	5	Q9V570	Q9v570 drosophila	893	6	3.1	795	11	O88469	O88469 rattus norv
821	6	3.1	683	10	Q8V2M4	Q8vzm4 arabidopsis	894	6	3.1	795	11	O88469	O88469 rattus norv
822	6	3.1	685	2	Q9F0N9	Q9f0n9 arosprilla	895	6	3.1	795	11	O921S3	O921s3 mus musculu
823	6	3.1	685	5	Q961G9	Q961g9 drosophila	896	6	3.1	796	5	Q17323	Q17323 caenorhabdi
824	6	3.1	687	5	Q95R14	Q95r14 drosophila	897	6	3.1	797	4	Q9UNN9	Q9unn9 homo sapien
825	6	3.1	689	3	Q06683	Q06683 saccharomyc	898	6	3.1	797	4	Q95267	Q95267 homo sapien
826	6	3.1	689	4	Q9HB58	Q9hb58 homo sapien	899	6	3.1	803	3	Q74402	Q74402 schizosacch
827	6	3.1	692	6	Q97746	Q97746 canis famil	900	6	3.1	806	11	Q9WVF4	Q9wvf4 mus musculu
828	6	3.1	693	10	Q9S1R0	Q9sir0 arabidopsis	901	6	3.1	809	10	Q8RWQ1	Q8rwq1 arabidopsis
829	6	3.1	695	2	Q9KK51	Q9kk51 streptococc	902	6	3.1	812	4	Q9BW43	Q9bw43 homo sapien
830	6	3.1	695	2	Q9KK36	Q9kk36 streptococc	903	6	3.1	813	16	Q97EC4	Q97ec4 clostridium
831	6	3.1	696	17	Q8TU16	Q8tui6 methanosarc	904	6	3.1	815	10	Q9SCN8	Q9scn8 arabidopsis
832	6	3.1	698	4	Q96MR6	Q96mr6 homo sapien	905	6	3.1	815	16	Q97EX5	Q97ex5 clostridium
833	6	3.1	698	5	Q8T2W9	Q8t2w9 trypanosoma	906	6	3.1	816	4	Q96G51	Q96g51 homo sapien
834	6	3.1	699	2	Q9KK42	Q9kk42 streptococc	907	6	3.1	816	12	O72163	O72163 feline astr
835	6	3.1	700	11	Q99K40	Q99k40 mus musculu	908	6	3.1	816	16	Q8XML8	Q8xml8 clostridium
836	6	3.1	702	2	Q93CM6	Q93cm6 bifidobacte	909	6	3.1	817	17	Q9U278	Q9u278 pyrococcus
837	6	3.1	703	5	Q61092	Q61092 trypanosoma	910	6	3.1	820	4	O60585	O60585 homo sapien
838	6	3.1	704	4	Q96QI3	Q96qi3 homo sapien	911	6	3.1	823	16	Q9AA14	Q9aa14 caulobacter
839	6	3.1	707	3	Q9P7G0	Q9p7g0 schizosacch	912	6	3.1	824	16	O8X904	O8x904 escherichia
840	6	3.1	707	6	Q28297	Q28297 canis famil	913	6	3.1	826	5	Q9XZQ1	Q9xzq1 caenorhabdi
841	6	3.1	707	10	Q94KD3	Q94kd3 arabidopsis	914	6	3.1	826	5	Q9XXH8	Q9xxh8 caenorhabdi
842	6	3.1	708	2	O82841	O82841 alteromonas	915	6	3.1	826	5	Q9XTX1	Q9xtx1 caenorhabdi
843	6	3.1	712	16	O8ZBG2	O8zbg2 yersinia pe	916	6	3.1	826	16	O80739	O80739 arabidopsis
844	6	3.1	713	5	Q19687	Q19687 caenorhabdi	917	6	3.1	826	16	Q92TC2	Q92tc2 rhizobium m
845	6	3.1	714	5	O45541	O45541 caenorhabdi	918	6	3.1	827	5	Q96029	Q96029 drosophila
846	6	3.1	717	5	O8T039	O8t039 drosophila	919	6	3.1	831	4	Q96MT3	Q96mt3 homo sapien
847	6	3.1	718	5	Q9V594	Q9v594 drosophila	920	6	3.1	832	5	Q9TZK8	Q9tzk8 caenorhabdi
848	6	3.1	720	2	Q9LRJ0	Q9lrj0 streptomyce	921	6	3.1	835	13	Q90Z06	Q90z06 xenopus lae
849	6	3.1	720	5	Q9VR99	Q9vr99 drosophila	922	6	3.1	843	10	Q9LZF6	Q9lzf6 arabidopsis
850	6	3.1	722	16	Q926X5	Q926x5 listeria in	923	6	3.1	847	16	Q9RRN5	Q9rrn5 deinococcus
851	6	3.1	724	10	Q9LW21	Q9lw21 arabidopsis	924	6	3.1	850	5	Q95TP1	Q95tp1 drosophila
852	6	3.1	727	17	Q9HH14	Q9hh14 halobacteri	925	6	3.1	852	10	Q9ZR08	Q9zr08 arabidopsis
853	6	3.1	728	10	Q9M9S4	Q9m9s4 arabidopsis	926	6	3.1	853	10	Q9FHJ6	Q9fhj6 arabidopsis
854	6	3.1	730	5	Q95QN2	Q95qn2 caenorhabdi	927	6	3.1	857	10	Q9ASK9	Q9ask9 oryza sativ
855	6	3.1	732	5	Q18367	Q18367 caenorhabdi	928	6	3.1	859	3	Q12102	Q12102 saccharomyc
856	6	3.1	733	10	Q9FK08	Q9fk08 arabidopsis	929	6	3.1	861	10	Q9SDE7	Q9sde7 oryza sativ
857	6	3.1	733	11	Q9D2C0	Q9d2c0 mus musculu	930	6	3.1	862	10	Q43191	Q43191 solanum tub
858	6	3.1	735	13	Q9DP09	Q9ddg9 xenopus lae	931	6	3.1	863	5	Q27267	Q27267 caenorhabdi
859	6	3.1	736	11	Q61164	Q61164 mus musculu	932	6	3.1	863	3	Q8X0P4	Q8x0p4 neurospora
860	6	3.1	737	11	Q9WVF3	Q9wvf3 mus musculu	933	6	3.1	866	13	Q90WV2	Q90wv2 xenopus lae
861	6	3.1	737	11	Q9R1D1	Q9r1d1 rattus norv	934	6	3.1	871	3	O74653	O74653 schizosacch
862	6	3.1	739	10	Q9LWY3	Q9lw3 oryza sativ	935	6	3.1	876	11	Q9WU62	Q9wu62 mus musculu
863	6	3.1	742	5	O9XZQ2	Q9xzq2 caenorhabdi	936	6	3.1	881	16	Q8UJA1	Q8uja1 agrobacteri
864	6	3.1	742	5	Q9XXH9	Q9xxh9 caenorhabdi	937	6	3.1	884	16	Q98E04	Q98e04 rhizobium l
865	6	3.1	743	2	Q8VPJ5	Q8vpj5 clostridium	938	6	3.1	892	5	Q9U299	Q9u299 caenorhabdi
866	6	3.1	743	3	Q9C2D1	Q9c2d1 neurospora	939	6	3.1	893	2	Q93V06	Q93v06 eubacterium
867	6	3.1	743	5	Q9NBV5	Q9nbv5 drosophila	940	6	3.1	897	11	O70495	O70495 mus musculu
868	6	3.1	749	5	Q17768	Q17768 caenorhabdi	941	6	3.1	900	10	O23454	O23454 arabidopsis
869	6	3.1	752	5	O21026	Q21026 caenorhabdi	942	6	3.1	901	12	Q8V2F7	Q8v2f7 bean leafro
870	6	3.1	752	5	O8SUP8	Q8sup8 encephalito	943	6	3.1	904	5	O45544	O45544 caenorhabdi
871	6	3.1	752	11	O55035	O55035 rattus norv	944	6	3.1	904	5	Q9W0F5	Q9w0f5 drosophila
872	6	3.1	753	5	O8T2G9	Q8t2g9 dictyosteli	945	6	3.1	911	17	Q975L8	Q975l8 sulfolobus
873	6	3.1	753	16	Q9K3L4	Q9k3l4 streptomyce	946	6	3.1	915	4	Q13827	Q13827 homo sapien
874	6	3.1	754	4	Q13427	Q13427 homo sapien	947	6	3.1	919	4	Q9NQS7	Q9nqs7 homo sapien
875	6	3.1	754	5	O8SXJ1	Q8sxj1 drosophila	948	6	3.1	923	10	O49003	O49003 avena sativ
876	6	3.1	755	6	Q8WP32	Q8wp32 macaca fasc	949	6	3.1	925	10	Q9SIX4	Q9sia4 arabidopsis
877	6	3.1	758	5	Q24086	Q24086 drosophila	950	6	3.1	927	10	O49004	O49004 avena sativ
878	6	3.1	759	5	Q9VWZ0	Q9vvz0 drosophila	951	6	3.1	938	5	Q9V3X5	Q9v3x5 drosophila
879	6	3.1	764	4	Q9P2H3	Q9p2h3 homo sapien	952	6	3.1	941	16	O83382	O83382 treponema p
880	6	3.1	765	5	O21478	Q21478 caenorhabdi	953	6	3.1	945	17	Q974N4	Q974n4 sulfolobus
881	6	3.1	768	16	O8UDA4	O8uda4 agrobacteri	954	6	3.1	950	13	Q9YHC9	Q9yhc9 xenopus lae
882	6	3.1	773	10	O22178	O22178 arabidopsis	955	6	3.1	950	16	Q9HXH0	Q9hxo0 pseudomonas
883	6	3.1	774	3	Q9HEC5	Q9hec5 neurospora	956	6	3.1	958	4	Q8WWI2	Q8wwi2 homo sapien
884	6	3.1	775	12	Q8QY62	Q8qy62 dengue viru	957	6	3.1	959	5	Q9VNI5	Q9vni5 drosophila
885	6	3.1	776	16	Q98A73	Q98a73 rhizobium l	958	6	3.1	960	5	Q9VY25	Q9vy25 drosophila
886	6	3.1	779	16	Q92HZ1	Q92hz1 rickettsia	959	6	3.1	962	5	O20625	O20625 caenorhabdi
887	6	3.1	784	16	O8ZC67	O8zc67 yersinia pe	960	6	3.1	964	5	Q8T856	Q8t856 dictyosteli
888	6	3.1	787	4	Q9UPY3	Q9upt3 homo sapien	961	6	3.1	967	17	Q8U2E6	Q8u2e6 pyrococcus
889	6	3.1	787	4	Q8TC42	Q8tc42 homo sapien	962	6	3.1	969	17	O28863	O28863 archaeoglob
890	6	3.1	787	4	Q8TC27	Q8tc27 homo sapien	963	6	3.1	970	5	Q9VAW1	Q9vaw1 drosophila
891	6	3.1	787	5	Q9VA09	Q9va09 drosophila	964	6	3.1	976	3	Q9C2H8	Q9c2h8 neurospora
892	6	3.1	791	16	Q98QC9	Q98qc9 mycoplasma	965	6	3.1	976	13	Q90ZN9	Q90zn9 brachydanio

674	6	3.1	539	4	Q9HCT8	Q9hct8 homo sapien	747	6	3.1	600	16	Q9HWY8	Q9hwy8 pseudomonas
675	6	3.1	541	10	Q94IJ3	Q94ij3 zea mays (m	748	6	3.1	601	5	O62185	O62185 caenorhabdi
676	6	3.1	542	4	Q9HA87	Q9ha87 homo sapien	749	6	3.1	601	10	Q9M9G3	Q9m9g3 arabidopsis
677	6	3.1	542	4	Q9H9J9	Q9h9j9 homo sapien	750	6	3.1	601	10	Q8VYT2	Q8vyt2 arabidopsis
678	6	3.1	542	16	Q50025	Q50025 mycobacteri	751	6	3.1	606	10	Q9FIW8	Q9fiw8 arabidopsis
679	6	3.1	545	4	Q9NXU7	Q9nxu7 homo sapien	752	6	3.1	607	10	Q9MAQ4	Q9maq4 arabidopsis
680	6	3.1	545	16	Q8WPL8	Q8wpl8 anabaena sp	753	6	3.1	613	10	Q9LQ94	Q9lq94 arabidopsis
681	6	3.1	547	4	Q8WUZ6	Q8wuz6 homo sapien	754	6	3.1	615	10	Q94F62	Q94f62 arabidopsis
682	6	3.1	547	5	Q962D4	Q962d4 ancyllostoma	755	6	3.1	617	5	O61395	O61395 caenorhabdi
683	6	3.1	548	2	O54414	O54414 acinetobact	756	6	3.1	617	10	Q9FIP1	Q9fip1 arabidopsis
684	6	3.1	548	13	Q9PTY5	Q9pty5 xenopus lae	757	6	3.1	618	5	O61394	O61394 caenorhabdi
685	6	3.1	550	4	Q9HCG0	Q9hcg0 homo sapien	758	6	3.1	619	5	Q9WID8	Q9wid8 drosophila
686	6	3.1	551	5	Q95TL9	Q95tl9 drosophila	759	6	3.1	619	11	Q91YS4	Q91ys4 mus musculu
687	6	3.1	552	4	Q9BR39	Q9br39 homo sapien	760	6	3.1	619	16	O67874	O67874 aquifex aeo
688	6	3.1	552	10	Q9C823	Q9c823 arabidopsis	761	6	3.1	620	8	Q99973	Q99973 porphyra pu
689	6	3.1	553	10	Q9C823	Q9c823 arabidopsis	762	6	3.1	620	8	Q99973	Q99973 porphyra pu
690	6	3.1	553	10	Q9C823	Q9c823 arabidopsis	763	6	3.1	620	8	Q99973	Q99973 porphyra pu
691	6	3.1	554	3	O74721	O74721 kluyveromyc	764	6	3.1	621	5	Q9VY09	Q9vy09 drosophila
692	6	3.1	554	5	O16351	O16351 caenorhabdi	765	6	3.1	621	10	Q948Y0	Q948y0 glycine max
693	6	3.1	556	5	Q8T434	Q8t434 drosophila	766	6	3.1	621	16	O66708	O66708 aquifex aeo
694	6	3.1	556	10	Q8RYU7	Q8ryu7 oryza sativ	767	6	3.1	622	5	Q95TY0	Q95ty0 drosophila
695	6	3.1	556	10	Q8RYU7	Q8ryu7 oryza sativ	768	6	3.1	622	10	Q93W70	Q93w70 zea mays (m
696	6	3.1	557	16	Q9XOV0	Q9xov0 thermotoga	769	6	3.1	625	10	Q94AG2	Q94ag2 arabidopsis
697	6	3.1	558	5	Q95R80	Q95r80 drosophila	770	6	3.1	625	11	O88991	O88991 mus musculu
698	6	3.1	558	5	Q9VCJ5	Q9vcj5 drosophila	771	6	3.1	626	10	Q94IJ5	Q94ij5 zea mays (m
699	6	3.1	558	11	Q91ZR0	Q91zr0 mus musculu	772	6	3.1	626	10	Q94IJ5	Q94ij5 zea mays (m
700	6	3.1	558	17	Q8TRW0	Q8trw0 methanopyru	773	6	3.1	627	11	Q922G0	Q922g0 mus musculu
701	6	3.1	559	10	Q9FZP9	Q9fzp9 glycine max	774	6	3.1	627	11	Q99NC9	Q99nc9 mus musculu
702	6	3.1	559	10	O8S3A6	O8s3a6 lactuca sat	775	6	3.1	627	11	Q922G0	Q922g0 mus musculu
703	6	3.1	560	3	O94667	O94667 schizosacch	776	6	3.1	627	11	Q9QZL3	Q9qzl3 mus musculu
704	6	3.1	560	10	Q94HQ4	Q94hq4 oryza sativ	777	6	3.1	627	16	O8Z0T5	O8z0t5 salmonella
705	6	3.1	561	5	O8SZS6	O8szs6 drosophila	778	6	3.1	628	10	Q9XIC7	Q9xic7 arabidopsis
706	6	3.1	561	10	Q9C7A9	Q9c7a9 arabidopsis	779	6	3.1	628	10	Q94F63	Q94f63 arabidopsis
707	6	3.1	562	5	Q9U345	Q9u345 caenorhabdi	780	6	3.1	629	3	O8XQ05	O8xq05 neospora
708	6	3.1	562	5	O61396	O61396 caenorhabdi	781	6	3.1	629	3	O8XQ05	O8xq05 neospora
709	6	3.1	563	5	O44136	O44136 caenorhabdi	782	6	3.1	631	16	O8XZJ4	O8xzj4 raistonia s
710	6	3.1	564	5	P91218	P91218 caenorhabdi	783	6	3.1	632	10	Q93ZS4	Q93zs4 arabidopsis
711	6	3.1	566	1	O73937	O73937 acidianus a	784	6	3.1	634	16	O8R715	O8r715 thermoanaer
712	6	3.1	566	16	Q9AKQ7	Q9akq7 rhizobium m	785	6	3.1	637	10	Q94DY8	Q94dy8 oryza sativ
713	6	3.1	566	17	Q9C4L9	Q9c4l9 sulfolobus	786	6	3.1	639	16	Q8YF90	Q8yf90 bruceella me
714	6	3.1	567	16	Q98BV2	Q98bv2 rhizobium l	787	6	3.1	639	17	O8ZT28	O8zt28 pyrobaculum
715	6	3.1	568	5	O9U344	Q9u344 caenorhabdi	788	6	3.1	639	17	O9UXE8	O9uxe8 sulfolobus
716	6	3.1	568	5	O02147	O02147 caenorhabdi	789	6	3.1	640	4	Q96PP9	Q96pp9 homo sapien
717	6	3.1	569	16	O8XVH4	O8xvh4 raistonia s	790	6	3.1	640	5	Q9VBR1	Q9vbr1 drosophila
718	6	3.1	570	10	Q9FLD0	Q9fld0 arabidopsis	791	6	3.1	640	16	Q9JZ98	Q9jz98 neisseria m
719	6	3.1	570	16	Q9PIC5	Q9pic5 campylobact	792	6	3.1	641	2	Q93EP2	Q93ep2 photorhabdu
720	6	3.1	570	17	Q8THH1	Q8thh1 methanosarc	793	6	3.1	645	17	O8TZP8	O8tzp8 pyrococcus
721	6	3.1	571	10	Q9SUW9	Q9siw9 arabidopsis	794	6	3.1	647	4	Q9Y4D2	Q9y4d2 homo sapien
722	6	3.1	571	13	O57604	O57604 gallus gall	795	6	3.1	649	10	Q949N7	Q949n7 arabidopsis
723	6	3.1	572	10	Q9LHB3	Q9lhb3 arabidopsis	796	6	3.1	649	16	O83749	O83749 treponema p
724	6	3.1	573	5	O8T022	O8t022 drosophila	797	6	3.1	650	2	O8RR67	O8rr67 actinobacil
725	6	3.1	573	10	Q9SIB0	Q9sib0 arabidopsis	798	6	3.1	650	5	O19878	O19878 caenorhabdi
726	6	3.1	573	10	Q9C7U0	Q9c7u0 arabidopsis	799	6	3.1	652	12	O85056	O85056 atkinsonell
727	6	3.1	574	10	Q9SZG9	Q9szg9 arabidopsis	800	6	3.1	653	2	O34097	O34097 streptococc
728	6	3.1	577	5	Q967G8	Q967g8 nippostrong	801	6	3.1	653	3	Q12171	Q12171 saccharomyc
729	6	3.1	577	10	Q9SW08	Q9sw08 arabidopsis	802	6	3.1	656	17	O8TWM1	O8twm1 methanopyru
730	6	3.1	577	16	O8XTI9	O8xti9 raistonia s	803	6	3.1	657	11	Q9EPG7	Q9epg7 mus musculu
731	6	3.1	578	5	Q9W324	Q9w324 drosophila	804	6	3.1	660	5	Q9VY90	Q9vy90 drosophila
732	6	3.1	580	10	Q93ZS9	Q93zs9 arabidopsis	805	6	3.1	661	5	O8T431	O8t431 drosophila
733	6	3.1	580	16	O8VYQ0	O8vyq0 anabaena sp	806	6	3.1	661	5	O9W245	O9w245 drosophila
734	6	3.1	581	10	Q9SVV9	Q9svv9 arabidopsis	807	6	3.1	661	17	Q96XA4	Q96xa4 sulfolobus
735	6	3.1	585	16	Q9KR47	Q9kr47 vibrio chol	808	6	3.1	662	10	Q9SKB8	Q9skb8 arabidopsis
736	6	3.1	586	2	Q9FDT7	Q9fdt7 microcystis	809	6	3.1	665	5	O9VXT5	O9vxt5 drosophila
737	6	3.1	586	10	Q9LY47	Q9ly47 arabidopsis	810	6	3.1	666	16	O8RXC3	O8rcx3 arabidopsis
738	6	3.1	588	10	O22717	O22717 arabidopsis	811	6	3.1	666	16	O8RC42	O8rc42 thermoanaer
739	6	3.1	589	2	Q9LBP2	Q9lbp2 haemophilus	812	6	3.1	667	11	O9R1D9	O9r1d9 mus musculu
740	6	3.1	589	10	Q9FLX5	Q9flx5 arabidopsis	813	6	3.1	668	12	O39328	O39328 feline calli
741	6	3.1	591	10	Q8S665	Q8s665 oryza sativ	814	6	3.1	668	12	O39329	O39329 feline calli
742	6	3.1	592	10	Q9LYW0	Q9lyw0 arabidopsis	815	6	3.1	673	5	O23706	O23706 carlupus ma
743	6	3.1	592	10	O04096	O04096 arabidopsis	816	6	3.1	674	5	O960F8	O960f8 drosophila
744	6	3.1	597	5	O17562	O17562 caenorhabdi	817	6	3.1	677	3	O74757	O74757 schizosacch
745	6	3.1	598	2	O9A141	O9a141 burkholderi	818	6	3.1	677	17	Q8TQS7	Q8tqs7 methanosarc
746	6	3.1	599	10	Q9U099	Q9u099 arabidopsis	819	6	3.1	678	2	Q93NP3	Q93np3 enterococcu

528	6	3.1	410	16	Q9WXA8	Q9WYA8 thermotoga	601	6	3.1	480	2	Q9LAX3	Q9LAX3 streptococc
529	6	3.1	411	4	Q9BT08	Q9BT08 homo sapien	602	6	3.1	481	2	Q9LAX5	Q9LAX5 streptococc
530	6	3.1	411	5	Q22718	Q22718 caenorhabdi	603	6	3.1	482	9	Q8SD07	Q8SD07 pseudomonas
531	6	3.1	411	10	Q949C6	Q949C6 oryza sativ	604	6	3.1	485	10	Q9ZQ07	Q9ZQ07 arabidopsis
532	6	3.1	412	13	Q9PUB0	Q9PUB0 coturnix co	605	6	3.1	485	12	Q84435	Q84435 paramecium
533	6	3.1	413	5	Q9VJPI	Q9VJPI rattus norv	606	6	3.1	487	16	Q69924	Q69924 streptomyce
534	6	3.1	413	11	Q35141	Q35141 rattus norv	607	6	3.1	490	17	Q97216	Q97216 sulfolobus
535	6	3.1	413	11	Q70407	Q70407 mus musculu	608	6	3.1	491	10	Q94455	Q94455 arabidopsis
536	6	3.1	413	13	Q9PUB1	Q9PUB1 anolis caro	609	6	3.1	492	5	Q18989	Q18989 caenorhabdi
537	6	3.1	416	16	Q8YRVO	Q8YRVO anabaena sp	610	6	3.1	492	17	Q97Y79	Q97Y79 sulfolobus
538	6	3.1	417	4	Q96N00	Q96N00 homo sapien	611	6	3.1	493	16	Q8YBP8	Q8YBP8 anabaena sp
539	6	3.1	418	4	Q8WWM0	Q8WWM0 homo sapien	612	6	3.1	494	2	Q9WXE7	Q9WXE7 prevotella
540	6	3.1	418	10	Q9LJI8	Q9LJI8 arabidopsis	613	6	3.1	494	11	Q9JK17	Q9JK17 rattus norv
541	6	3.1	418	17	Q8TW51	Q8TW51 methanosarc	614	6	3.1	495	12	Q66459	Q66459 dengue viru
542	6	3.1	421	5	Q9WH79	Q9WH79 drosophila	615	6	3.1	495	12	Q66459	Q66459 dengue viru
543	6	3.1	422	10	Q9LU40	Q9LU40 arabidopsis	616	6	3.1	495	12	Q66460	Q66460 dengue viru
544	6	3.1	423	10	Q9FY60	Q9FY60 arabidopsis	617	6	3.1	495	12	Q66461	Q66461 dengue viru
545	6	3.1	423	10	Q94CG3	Q94CG3 arabidopsis	618	6	3.1	495	12	Q66462	Q66462 dengue viru
546	6	3.1	424	10	Q64420	Q64420 cyanidium c	619	6	3.1	495	12	Q66463	Q66463 dengue viru
547	6	3.1	424	10	Q64419	Q64419 cyanidium c	620	6	3.1	495	12	Q66464	Q66464 dengue viru
548	6	3.1	425	8	Q23890	Q23890 dictyosteli	621	6	3.1	495	12	Q9QCX0	Q9QCX0 dengue viru
549	6	3.1	425	11	Q9CQ38	Q9CQ38 mus musculu	622	6	3.1	495	12	Q9QCW6	Q9QCW6 dengue viru
550	6	3.1	428	2	Q9ANY6	Q9ANY6 enterococcu	623	6	3.1	495	12	Q9QCW5	Q9QCW5 dengue viru
551	6	3.1	430	5	Q95VA6	Q95VA6 caenorhabdi	624	6	3.1	495	12	Q9QCW4	Q9QCW4 dengue viru
552	6	3.1	431	10	Q94E25	Q94E25 oryza sativ	625	6	3.1	495	12	Q9QCW2	Q9QCW2 dengue viru
553	6	3.1	432	5	Q9VUX4	Q9VUX4 drosophila	626	6	3.1	495	12	Q90203	Q90203 dengue viru
554	6	3.1	434	6	Q9GKV7	Q9GKV7 macaca fasc	627	6	3.1	496	2	Q52249	Q52249 enterococcu
555	6	3.1	435	10	Q9FYM0	Q9FYM0 arabidopsis	628	6	3.1	496	2	Q93V07	Q93V07 enterococcu
556	6	3.1	435	10	Q94Y10	Q94Y10 arabidopsis	629	6	3.1	496	10	Q9SS55	Q9SS55 arabidopsis
557	6	3.1	437	17	Q8TLL2	Q8TLL2 methanosarc	630	6	3.1	497	16	Q9X293	Q9X293 thermotoga
558	6	3.1	438	10	Q9M050	Q9M050 arabidopsis	631	6	3.1	498	16	Q9K994	Q9K994 bacillus ha
559	6	3.1	439	11	Q91W08	Q91W08 mus musculu	632	6	3.1	501	5	Q20179	Q20179 caenorhabdi
560	6	3.1	443	4	Q96F64	Q96F64 homo sapien	633	6	3.1	502	5	Q20280	Q20280 caenorhabdi
561	6	3.1	443	16	Q97NQ5	Q97NQ5 streptococc	634	6	3.1	502	5	Q8WQB8	Q8WQB8 caenorhabdi
562	6	3.1	444	10	Q9C8C3	Q9C8C3 arabidopsis	635	6	3.1	503	2	Q93TT3	Q93TT3 vibrio chol
563	6	3.1	444	10	Q9C8C2	Q9C8C2 arabidopsis	636	6	3.1	503	4	Q8TAJ1	Q8TAJ1 homo sapien
564	6	3.1	445	2	Q68574	Q68574 streptococc	637	6	3.1	504	5	Q16352	Q16352 caenorhabdi
565	6	3.1	448	5	Q9NKE0	Q9NKE0 drosophila	638	6	3.1	504	8	Q95GT7	Q95GT7 nepenthes d
566	6	3.1	449	17	Q8TP09	Q8TP09 methanosarc	639	6	3.1	506	17	Q8TMY4	Q8TMY4 methanosarc
567	6	3.1	450	2	Q93H56	Q93H56 streptomyce	640	6	3.1	508	4	Q8WXA9	Q8WXA9 homo sapien
568	6	3.1	451	2	Q9LAB9	Q9LAB9 pseudomonas	641	6	3.1	509	10	Q8VZR6	Q8VZR6 arabidopsis
569	6	3.1	451	5	Q20648	Q20648 caenorhabdi	642	6	3.1	510	12	Q91GE8	Q91GE8 epiphyas po
570	6	3.1	452	17	Q97V68	Q97V68 sulfolobus	643	6	3.1	512	3	Q94529	Q94529 schizosacch
571	6	3.1	453	5	Q8SZA0	Q8SZA0 drosophila	644	6	3.1	512	10	Q9SV85	Q9SV85 arabidopsis
572	6	3.1	454	16	Q97SF8	Q97SF8 streptococc	645	6	3.1	514	5	Q8SUD0	Q8SUD0 encephalito
573	6	3.1	456	11	Q9JLL9	Q9JLL9 mus musculu	646	6	3.1	515	5	Q8T8Z5	Q8T8Z5 drosophila
574	6	3.1	457	3	Q43087	Q43087 schizosacch	647	6	3.1	516	3	Q96VU3	Q96VU3 leishmania e
575	6	3.1	457	5	Q02096	Q02096 caenorhabdi	648	6	3.1	519	10	Q41714	Q41714 welwitschia
576	6	3.1	457	16	Q92VC3	Q92VC3 rhizobium m	649	6	3.1	520	10	Q9SKG5	Q9SKG5 arabidopsis
577	6	3.1	460	10	Q9C8H5	Q9C8H5 arabidopsis	650	6	3.1	520	10	Q9SRL6	Q9SRL6 arabidopsis
578	6	3.1	461	2	Q9LAX6	Q9LAX6 streptococc	651	6	3.1	520	10	Q9SUD2	Q9SUD2 arabidopsis
579	6	3.1	461	16	Q8RB46	Q8RB46 thermoanaer	652	6	3.1	521	10	Q22848	Q22848 arabidopsis
580	6	3.1	463	5	Q9VXJ8	Q9VXJ8 drosophila	653	6	3.1	522	5	Q9VNZ4	Q9VNZ4 drosophila
581	6	3.1	464	10	Q8S3B7	Q8S3B7 glycoline max	654	6	3.1	522	5	Q8T9G3	Q8T9G3 drosophila
582	6	3.1	464	16	Q8YOG4	Q8YOG4 anabaena sp	655	6	3.1	523	3	Q9Y723	Q9Y723 irpep lacte
583	6	3.1	465	10	Q8S3B8	Q8S3B8 glycine max	656	6	3.1	523	10	Q9SZC0	Q9SZC0 arabidopsis
584	6	3.1	465	16	Q935I8	Q935I8 salmonella	657	6	3.1	524	10	Q9SKG4	Q9SKG4 arabidopsis
585	6	3.1	466	3	Q9USN2	Q9USN2 schizosacch	658	6	3.1	525	3	Q12621	Q12621 humicola gr
586	6	3.1	466	4	Q9H9U2	Q9H9U2 homo sapien	659	6	3.1	525	16	Q8RG92	Q8RG92 fusobacteri
587	6	3.1	468	2	Q9LSJ4	Q9LSJ4 salmonella	660	6	3.1	527	10	Q9LY45	Q9LY45 arabidopsis
588	6	3.1	469	16	Q9KMK32	Q9KMK32 vibrio chol	661	6	3.1	527	16	Q98NA7	Q98NA7 rhizobium l
589	6	3.1	470	2	Q8VU05	Q8VU05 streptococc	662	6	3.1	528	4	Q9H6U3	Q9H6U3 homo sapien
590	6	3.1	470	2	Q9AHE1	Q9AHE1 streptococc	663	6	3.1	530	5	Q9BKR8	Q9BKR8 caenorhabdi
591	6	3.1	470	2	Q8RLP8	Q8RLP8 streptococc	664	6	3.1	531	5	Q8T8A7	Q8T8A7 clona intes
592	6	3.1	471	12	Q9ELZ8	Q9ELZ8 cercopitheci	665	6	3.1	531	6	Q9XS55	Q9XS55 ovine aries
593	6	3.1	472	5	Q20101	Q20101 caenorhabdi	666	6	3.1	532	6	Q9XS56	Q9XS56 ovine aries
594	6	3.1	472	5	Q01631	Q01631 caenorhabdi	667	6	3.1	533	4	Q15317	Q15317 homo sapien
595	6	3.1	473	16	Q8XVP1	Q8XVP1 ralstonia s	668	6	3.1	533	10	Q8RZ85	Q8RZ85 oryza sativ
596	6	3.1	476	2	Q9ZG91	Q9ZG91 pseudomonas	669	6	3.1	533	11	Q9D483	Q9D483 mus musculu
597	6	3.1	477	4	Q96JN6	Q96JN6 homo sapien	670	6	3.1	534	4	Q9Y3R6	Q9Y3R6 mus sapien
598	6	3.1	477	17	Q8TQW2	Q8TQW2 methanosarc	671	6	3.1	534	4	Q9BU14	Q9BU14 homo sapien
599	6	3.1	478	5	Q9NKE7	Q9NKE7 drosophila	672	6	3.1	536	3	Q9Y895	Q9Y895 volvariella
600	6	3.1	479	2	Q9LAX2	Q9LAX2 streptococc	673	6	3.1	538	2	Q9RNB6	Q9RNB6 microcystis

382	6	3.1	314	16	Q92EP8	Q92ep8 listeria mo	455	6	3.1	363	2	Q56612	Q56612 vibrio chol
383	6	3.1	315	4	Q8WVH5	Q8wvh5 homo sapien	456	6	3.1	363	16	Q8UG63	Q8ug63 agrobacteri
384	6	3.1	315	10	Q9SW35	Q9sw35 arabidopsis	457	6	3.1	365	11	Q9UJ37	Q9uj37 mus musculu
385	6	3.1	317	16	Q9KNE7	Q9knf7 vibrio chol	458	6	3.1	367	11	Q9D5P8	Q9d5p8 mus musculu
386	6	3.1	317	16	Q97G43	Q97g43 clostridium	459	6	3.1	369	12	Q9DH64	Q9dh64 meleagrid h
387	6	3.1	318	16	Q97MV2	Q97mv2 clostridium	460	6	3.1	372	4	Q60529	Q60529 homo sapien
388	6	3.1	319	8	Q34805	Q34805 nansenula w	461	6	3.1	373	2	Q9F2H7	Q9f2h7 staphylococ
389	6	3.1	319	16	Q34966	Q34966 bacillus su	462	6	3.1	373	10	Q9LT76	Q9lt76 arabidopsis
390	6	3.1	321	17	Q8ZTA3	Q8zta3 pyrobaculum	463	6	3.1	374	16	Q9KFW8	Q9kfw8 bacillus ha
391	6	3.1	321	5	Q20711	Q20711 caenorhabdi	464	6	3.1	375	10	Q9SVA0	Q9sva0 arabidopsis
392	6	3.1	321	10	Q9XFF0	Q9xff0 avena sativ	465	6	3.1	376	8	Q9MSF2	Q9msf2 nepenthes d
393	6	3.1	321	10	Q23662	Q23662 arabidopsis	466	6	3.1	378	4	Q9UL09	Q9ul09 homo sapien
394	6	3.1	321	16	Q92VR8	Q92vr8 rhizobium m	467	6	3.1	378	10	Q9CA65	Q9cac65 arabidopsis
395	6	3.1	324	5	Q9VWR0	Q9vwr0 drosophila	468	6	3.1	379	3	Q9CA41	Q9cac41 alternaria
396	6	3.1	325	16	Q92EQ5	Q92eq5 listeria in	469	6	3.1	379	3	Q9C4A0	Q9cac40 alternaria
397	6	3.1	325	16	Q8Y9V2	Q8y9v2 listeria mo	470	6	3.1	379	17	Q58234	Q58234 pyrococcus
398	6	3.1	325	16	Q8R9V6	Q8r9v6 thermoanaer	471	6	3.1	379	17	Q97X11	Q97x11 sulfolobus
399	6	3.1	328	4	Q96E84	Q96e84 homo sapien	472	6	3.1	379	17	Q8U3V2	Q8u3v2 pyrococcus
400	6	3.1	328	11	Q9CSU9	Q9csu9 mus musculu	473	6	3.1	380	4	Q9H894	Q9h894 homo sapien
401	6	3.1	328	16	Q9WZB6	Q9wze6 thermotoga	474	6	3.1	382	17	Q9UYG2	Q9uyg2 pyrococcus
402	6	3.1	329	16	Q8U758	Q8u758 agrobacteri	475	6	3.1	383	16	Q25795	Q25795 helicobacte
403	6	3.1	331	3	Q00893	Q00893 collettetric	476	6	3.1	383	16	Q9ZK33	Q9zk33 helicobacte
404	6	3.1	331	9	Q64366	Q64366 bacterioph	477	6	3.1	385	1	Q59569	Q59569 methanobact
405	6	3.1	332	4	Q60530	Q60530 homo sapien	478	6	3.1	385	4	Q9GZS6	Q9gzs6 homo sapien
406	6	3.1	335	16	Q9Z778	Q9z778 chlamydia p	479	6	3.1	385	4	Q9Y306	Q9y306 homo sapien
407	6	3.1	336	2	Q9ZGL4	Q9zgl4 leptospira	480	6	3.1	385	4	Q9BRX2	Q9brx2 homo sapien
408	6	3.1	336	2	Q9S4H0	Q9s4h0 leptospira	481	6	3.1	385	11	Q91U22	Q91uz2 mus musculu
409	6	3.1	336	3	Q8TFK1	Q8tfk1 debaryomyce	482	6	3.1	386	5	Q36099	Q36099 theileria p
410	6	3.1	338	10	Q9M2D6	Q9m2d6 arabidopsis	483	6	3.1	387	5	Q9V4V4	Q9v4v4 drosophila
411	6	3.1	339	10	Q9FQ77	Q9fq77 zea mays (m	484	6	3.1	388	5	Q9XTC8	Q9xtc8 caenorhabdi
412	6	3.1	341	16	Q9XNL8	Q9xn18 rhizobium l	485	6	3.1	389	5	Q97253	Q97253 plasmodium
413	6	3.1	342	16	Q9XNS7	Q9xn57 clostridium	486	6	3.1	389	16	Q86581	Q86581 streptomyce
414	6	3.1	343	2	Q31098	Q31098 flavobacter	487	6	3.1	391	5	Q9XVL6	Q9xvl6 caenorhabdi
415	6	3.1	343	5	Q18070	Q18070 caenorhabdi	488	6	3.1	391	10	Q9ZSL4	Q9zsl4 arabidopsis
416	6	3.1	344	11	Q89037	Q89037 rattus norv	489	6	3.1	392	5	Q9VYI8	Q9vyi8 drosophila
417	6	3.1	344	16	Q34829	Q34829 bacillus su	490	6	3.1	393	11	Q8R4U1	Q8r4ul mus musculu
418	6	3.1	345	3	Q74906	Q74906 schizosacch	491	6	3.1	394	12	Q9Q1Z4	Q9qlz4 rotavirus c
419	6	3.1	345	16	Q99UY5	Q99uy5 staphylococ	492	6	3.1	395	10	Q9LUB8	Q9lub8 arabidopsis
420	6	3.1	346	10	Q9SYB5	Q9syb5 arabidopsis	493	6	3.1	395	10	Q9FR67	Q9fr67 cardamine p
421	6	3.1	348	2	Q9ANP0	Q9anp0 bradyrhizob	494	6	3.1	395	12	Q8QLI0	Q8qli0 mamestra co
422	6	3.1	348	16	Q92M44	Q92m44 rhizobium m	495	6	3.1	395	16	Q9KU92	Q9ku92 vibrio chol
423	6	3.1	349	16	Q8ZNK7	Q8znk7 salmonella	496	6	3.1	395	16	Q9CMQ8	Q9cmq8 pasteurilla
424	6	3.1	349	16	Q8Z594	Q8z594 salmonella	497	6	3.1	396	12	P88950	P88950 kaposi's sa
425	6	3.1	350	2	Q9RNL7	Q9rnl7 zymomonas m	498	6	3.1	397	10	Q9SFD1	Q9sfd1 arabidopsis
426	6	3.1	350	13	Q93565	Q93565 xenopus lae	499	6	3.1	399	2	Q9ZHA8	Q9zha8 streptococc
427	6	3.1	350	16	Q929G1	Q929g1 listeria in	500	6	3.1	399	5	Q9VHS8	Q9vhs8 drosophila
428	6	3.1	351	5	Q9N7S0	Q9n7s0 leishmania	501	6	3.1	399	5	Q8SXH3	Q8sxh3 drosophila
429	6	3.1	351	16	Q8U7V6	Q8u7v6 agrobacteri	502	6	3.1	399	16	Q915Q3	Q915q3 pseudomonas
430	6	3.1	352	10	Q9SVY7	Q9svy7 arabidopsis	503	6	3.1	399	16	Q97RU7	Q97ru7 streptococc
431	6	3.1	353	4	Q96QH5	Q96qh5 homo sapien	504	6	3.1	399	16	Q99SY4	Q99sy4 staphylococ
432	6	3.1	353	4	Q8WWR5	Q8wwr5 homo sapien	505	6	3.1	400	10	Q8S3I8	Q8s3i8 triticum ae
433	6	3.1	353	16	Q9RY70	Q9ry70 deinococcus	506	6	3.1	400	16	Q97KY6	Q97ky6 clostridium
434	6	3.1	353	16	Q8Y564	Q8y564 listeria mo	507	6	3.1	401	2	Q9LAZ2	Q9laz2 streptococc
435	6	3.1	354	17	Q976N6	Q976n6 sulfolobus	508	6	3.1	401	10	Q9SFD0	Q9sfd0 arabidopsis
436	6	3.1	354	4	Q9UGA4	Q9uga4 homo sapien	509	6	3.1	402	4	Q9NVB6	Q9nvb6 homo sapien
437	6	3.1	354	4	Q8TAN5	Q8tan5 homo sapien	510	6	3.1	402	4	Q9H7U5	Q9h7u5 homo sapien
438	6	3.1	354	5	Q8WP45	Q8wp45 halocynthia	511	6	3.1	402	5	Q95Q70	Q95q70 caenorhabdi
439	6	3.1	354	10	Q40067	Q40067 hordeum vul	512	6	3.1	403	5	Q8SVA8	Q8sva8 encephalito
440	6	3.1	354	17	Q28281	Q28281 archaeoglob	513	6	3.1	403	17	Q9YFN7	Q9yfn7 aeropyrum p
441	6	3.1	355	4	Q96C71	Q96c71 homo sapien	514	6	3.1	404	10	Q9C8C4	Q9c8c4 arabidopsis
442	6	3.1	355	5	Q17757	Q17757 caenorhabdi	515	6	3.1	404	16	Q9A3T0	Q9a3t0 caulobacter
443	6	3.1	355	10	Q9VZF4	Q9vzf4 arabidopsis	516	6	3.1	406	11	Q9JL35	Q9jl35 mus musculu
444	6	3.1	355	13	Q9W6A4	Q9w6a4 squalus aca	517	6	3.1	406	11	Q8VC71	Q8vc71 mus musculu
445	6	3.1	356	9	Q8SCJ7	Q8scj7 bacterioph	518	6	3.1	407	16	Q8X598	Q8x598 escherichia
446	6	3.1	357	5	Q9GRQ3	Q9grq3 leishmania	519	6	3.1	407	4	Q96GL7	Q96gl7 homo sapien
447	6	3.1	358	3	Q13754	Q13754 schizosacch	520	6	3.1	407	10	Q85879	Q85879 pisum sativ
448	6	3.1	359	2	Q9F329	Q9f329 staphylococ	521	6	3.1	407	16	Q9KUV6	Q9kuv6 vibrio chol
449	6	3.1	359	5	Q9W0Z5	Q9w0z5 drosophila	522	6	3.1	408	4	Q14977	Q14977 homo sapien
450	6	3.1	360	1	P96085	P96085 thermoplasma	523	6	3.1	408	10	Q94A52	Q94a52 arabidopsis
451	6	3.1	360	17	Q9HI56	Q9hi56 thermoplasma	524	6	3.1	409	4	Q9HA07	Q9ha07 homo sapien
452	6	3.1	361	16	Q9CGI4	Q9cqi4 lactococcus	525	6	3.1	409	4	Q9H0R7	Q9h0r7 homo sapien
453	6	3.1	362	16	Q9ZCH3	Q9zch3 lactococcus	526	6	3.1	409	4	Q96EQ1	Q96eq1 homo sapien
454	6	3.1	362	16	Q9ZGC4	Q9zgc4 rickettsia	527	6	3.1	409	4	Q8TB45	Q8tb45 homo sapien

236	6	3.1	227	17	Q82VD4	Q82vd4 pyrobaculum	309	6	3.1	265	4	Q8Wxf4	Q8wxf4 homo sapien
237	6	3.1	228	10	Q8S3K1	Q8s3k1 helianthus	310	6	3.1	265	5	Q9Gu88	Q9gu88 paramecium
238	6	3.1	228	10	Q8S3J9	Q8s3j9 helianthus	311	6	3.1	265	11	Q9CVG5	Q9cvg5 mus musculu
239	6	3.1	229	17	Q8ZUQ0	Q8zuq0 pyrobaculum	312	6	3.1	265	16	Q26015	Q26015 helicobacte
240	6	3.1	230	16	Q9CGH1	Q9cgh1 lactococcus	313	6	3.1	265	16	Q8ZJD8	Q8zjd8 helicobacte
241	6	3.1	231	3	Q93985	Q93985 neocallimas	314	6	3.1	265	16	Q8VJ03	Q8vj03 mycobacteri
242	6	3.1	233	2	Q9L568	Q9l568 streptococc	315	6	3.1	266	10	Q9LXy9	Q9lx99 arabidopsis
243	6	3.1	234	4	Q9BVI6	Q9bvi6 homo sapien	316	6	3.1	266	16	Q9JXH1	Q9jxh1 neisseria m
244	6	3.1	234	5	Q9XYW7	Q9xyw7 entodinium	317	6	3.1	267	16	Q8YF70	Q8yf70 brucella me
245	6	3.1	236	2	Q9L569	Q9l569 streptococc	318	6	3.1	268	5	Q9GU87	Q9gu87 paramecium
246	6	3.1	236	16	Q32238	Q32238 bacillus su	319	6	3.1	268	16	Q92TU6	Q92tu6 rhizobium m
247	6	3.1	237	17	Q8ZYL9	Q8zy19 pyrobaculum	320	6	3.1	270	17	Q26323	Q26323 methanobact
248	6	3.1	238	2	Q9RBP4	Q9rbp4 rhodococcus	321	6	3.1	270	16	Q8VX00	Q8vxx0 anaebaena sp
249	6	3.1	238	10	Q8S3K0	Q8s3k0 helianthus	322	6	3.1	272	16	Q9BR42	Q9br42 homo sapien
250	6	3.1	239	17	Q9Y0U6	Q9y0u6 pyrococcus	323	6	3.1	274	17	Q8TJV1	Q8tjv1 methanosarc
251	6	3.1	239	17	Q8U143	Q8u143 pyrococcus	324	6	3.1	276	2	Q8VVU1	Q8vvul staphylococ
252	6	3.1	241	16	Q8RGM2	Q8rgm2 fusobacteri	325	6	3.1	276	10	Q93Y10	Q93y10 arabidopsis
253	6	3.1	241	17	Q8TM92	Q8tm92 methanosarc	326	6	3.1	277	2	Q937M1	Q937m1 uncultured
254	6	3.1	242	11	Q9D6J2	Q9d6j2 mus musculu	327	6	3.1	277	2	Q937L6	Q937l6 uncultured
255	6	3.1	242	17	Q27260	Q27260 methanobact	328	6	3.1	277	2	Q932Y8	Q932y8 uncultured
256	6	3.1	243	2	Q9L567	Q9l567 streptococc	329	6	3.1	277	3	Q13660	Q13660 schizosacch
257	6	3.1	243	2	Q9L564	Q9l564 streptococc	330	6	3.1	277	5	Q8SS59	Q8ss59 encephalito
258	6	3.1	243	16	Q8Y3M0	Q8y3m0 listeria mo	331	6	3.1	279	16	Q9KCS3	Q9kcs3 bacillus ha
259	6	3.1	244	2	Q9L565	Q9l565 streptococc	332	6	3.1	280	2	Q9EVD3	Q9evd3 neisseria s
260	6	3.1	244	17	Q59140	Q59140 pyrococcus	333	6	3.1	280	2	Q93EK4	Q93ek4 neisseria m
261	6	3.1	245	5	Q8T9Z7	Q8t9z7 plasmidium	334	6	3.1	281	10	Q8S0I5	Q8s0i5 oryza sativ
262	6	3.1	246	2	Q45023	Q45023 borrelia bu	335	6	3.1	283	4	Q96S11	Q96s11 homo sapien
263	6	3.1	246	2	Q9L5B4	Q9l5b4 streptococc	336	6	3.1	283	6	Q9XS28	Q9xs28 cercopithe
264	6	3.1	246	2	Q8VS45	Q8vs45 borrelia bu	337	6	3.1	283	10	Q9SW52	Q9sw52 arabidopsis
265	6	3.1	246	2	Q8VS43	Q8vs43 borrelia bu	338	6	3.1	283	16	Q8ZQS9	Q8zqs9 salmonella
266	6	3.1	246	2	Q68233	Q68233 borrelia bu	339	6	3.1	284	5	Q16724	Q16724 caenorhabdi
267	6	3.1	247	2	Q9L566	Q9l566 streptococc	340	6	3.1	285	10	Q9FLE7	Q9fle7 arabidopsis
268	6	3.1	247	4	Q15262	Q15262 homo sapien	341	6	3.1	286	2	Q87477	Q87477 chlamydia t
269	6	3.1	247	16	Q8RAA1	Q8raa1 thermoanaer	342	6	3.1	286	5	Q94390	Q94390 caenorhabdi
270	6	3.1	248	4	Q14976	Q14976 homo sapien	343	6	3.1	286	17	Q9Y9J3	Q9y9j3 aeropyrum p
271	6	3.1	248	10	Q9X150	Q9xi50 arabidopsis	344	6	3.1	286	17	Q8TRC9	Q8trc9 methanosarc
272	6	3.1	248	16	Q98AL3	Q98al3 rhizobium l	345	6	3.1	287	2	Q9AIX3	Q9aix3 western x p
273	6	3.1	249	2	Q9S0B3	Q9s0b3 borrelia bu	346	6	3.1	288	10	Q82396	Q82396 arabidopsis
274	6	3.1	249	2	Q68238	Q68238 borrelia bu	347	6	3.1	291	4	Q9BZ75	Q9bz75 homo sapien
275	6	3.1	249	2	Q9FBA7	Q9fba7 borrelia he	348	6	3.1	291	4	Q96SB6	Q96sb6 homo sapien
276	6	3.1	249	2	Q9L575	Q9l575 streptococc	349	6	3.1	291	4	Q96EM3	Q96em3 homo sapien
277	6	3.1	249	2	Q9L570	Q9l570 streptococc	350	6	3.1	292	17	Q27819	Q27819 methanobact
278	6	3.1	249	2	Q8VS44	Q8vs44 borrelia bu	351	6	3.1	294	10	Q9S826	Q9s826 arabidopsis
279	6	3.1	251	5	Q45595	Q45595 caenorhabdi	352	6	3.1	296	5	Q17357	Q17357 caenorhabdi
280	6	3.1	252	16	Q50715	Q50715 borrelia bu	353	6	3.1	296	16	Q8YDX5	Q8ydx5 brucella me
281	6	3.1	252	17	Q97ZV7	Q97zv7 sulfolobus	354	6	3.1	297	2	Q8RWE1	Q8rme1 bacillus th
282	6	3.1	252	17	Q973S4	Q97zs4 sulfolobus	355	6	3.1	297	12	Q91RN3	Q9lrn3 rabies viru
283	6	3.1	252	17	Q96Y77	Q96y77 sulfolobus	356	6	3.1	297	12	Q91RL6	Q9lr16 rabies viru
284	6	3.1	253	2	Q9ZGC1	Q9zgc1 streptomyce	357	6	3.1	297	12	Q91RG3	Q9lrg3 rabies viru
285	6	3.1	253	5	Q9BII4	Q9bii4 paragonimus	358	6	3.1	298	4	Q9BRM4	Q9brm4 homo sapien
286	6	3.1	254	2	Q9L563	Q9l563 streptococc	359	6	3.1	298	16	Q50682	Q50682 borrelia bu
287	6	3.1	255	12	Q89459	Q89459 heliothis a	360	6	3.1	298	16	Q8RBG7	Q8rbg7 thermoanaer
288	6	3.1	256	11	Q9L595	Q9l595 streptococc	361	6	3.1	299	5	P91959	P91959 psammecinu
289	6	3.1	256	11	Q99J75	Q99jt5 mus musculu	362	6	3.1	299	16	Q98H90	Q98h90 rhizobium l
290	6	3.1	257	12	Q9E350	Q9e350 maize negro	363	6	3.1	299	17	Q9HQP8	Q9hq88 halobacteri
291	6	3.1	257	13	Q9PD91	Q9pd91 brachydanio	364	6	3.1	301	16	Q8RBH7	Q8rbh7 thermoanaer
292	6	3.1	258	5	Q16317	Q16317 caenorhabdi	365	6	3.1	302	16	Q8UKB7	Q8ukb7 agrobacteri
293	6	3.1	259	4	Q9NFI7	Q9nfi7 homo sapien	366	6	3.1	303	2	Q67986	Q67986 rhodococcus
294	6	3.1	261	4	Q96P17	Q96p17 homo sapien	367	6	3.1	303	5	Q18306	Q18306 caenorhabdi
295	6	3.1	261	4	Q8WXF0	Q8wxf0 homo sapien	368	6	3.1	303	5	O45673	O45673 caenorhabdi
296	6	3.1	261	4	Q8WW25	Q8ww25 homo sapien	369	6	3.1	304	5	Q23337	Q23337 caenorhabdi
297	6	3.1	261	16	Q9CJ60	Q9cj60 lactococcus	370	6	3.1	304	10	Q8S3F3	Q8s3f3 arabidopsis
298	6	3.1	262	4	Q75494	Q75494 homo sapien	371	6	3.1	305	10	Q93X11	Q93x11 panicum max
299	6	3.1	262	5	Q9VBK4	Q9vbk4 drosophila	372	6	3.1	305	17	Q9YDG7	Q9ydg7 aeropyrum p
300	6	3.1	262	11	Q9R000	Q9r0u0 mus musculu	373	6	3.1	306	16	Q9AA23	Q9aa23 caulobacter
301	6	3.1	262	11	Q88468	Q88468 mus musculu	374	6	3.1	307	10	Q9M2U1	Q9m2u1 arabidopsis
302	6	3.1	262	16	Q9JWG4	Q9jwg4 neisseria m	375	6	3.1	309	16	P73682	P73682 synecocyst
303	6	3.1	262	16	Q97HG9	Q97hg9 clostridium	376	6	3.1	310	17	Q9UXJ4	Q9uxj4 sulfolobus
304	6	3.1	263	12	Q86274	Q86274 avian rotav	377	6	3.1	312	10	Q9X176	Q9xi76 arabidopsis
305	6	3.1	263	17	Q97B90	Q97b90 thermoplas	378	6	3.1	312	17	Q97A20	Q97a20 thermoplasm
306	6	3.1	264	10	Q9XG04	Q9xgu4 arabidopsis	379	6	3.1	313	17	Q96YB6	Q96yb6 sulfolobus
307	6	3.1	265	2	Q87252	Q87252 lactococcus	380	6	3.1	314	10	Q9LST2	Q9lst2 arabidopsis
308	6	3.1	265	2	Q53186	Q53186 rhizobium s	381	6	3.1	314	11	Q9WU66	Q9wu66 mus musculu

90	6	3.1	55	16	Q8XM48	Q8xm48 clostridium	163	6	3.1	153	5	Q8WP53	Q8wp53 solenopsis
91	6	3.1	56	16	Q9PCT3	Q9pct3 xylella fas	164	6	3.1	153	13	Q90XE5	Q90xe5 xenopus lae
92	6	3.1	60	16	Q26313	Q26313 drosophila	165	6	3.1	156	5	O17325	O17325 caenorhabdi
93	6	3.1	61	5	O8STG3	O8stg3 echinococu	166	6	3.1	156	5	Q8WRP6	Q8wrp6 solenopsis
94	6	3.1	61	17	Q978C2	Q978c2 thermoplasm	167	6	3.1	158	16	Q9PKX4	Q9pkx4 chlamydia m
95	6	3.1	68	4	O95048	O95048 homo sapien	168	6	3.1	160	4	Q9P0C0	Q9p0c0 homo sapien
96	6	3.1	71	17	Q8TQ60	Q8tqe0 methanosarc	169	6	3.1	161	4	Q9NZA7	Q9nza7 homo sapien
97	6	3.1	74	12	O39651	O39651 dengue viru	170	6	3.1	166	4	Q9HA52	Q9ha52 homo sapien
98	6	3.1	74	12	O39652	O39652 dengue viru	171	6	3.1	166	10	O40652	O40652 oryza sativ
99	6	3.1	76	10	Q9X1K9	Q9xik9 arabidopsis	172	6	3.1	167	5	Q9TZH3	Q9tzh3 caenorhabdi
100	6	3.1	79	5	Q9NE61	Q9ne61 leishmania	173	6	3.1	168	4	O16043	O16043 homo sapien
101	6	3.1	82	13	Q8QG36	Q8qgy6 fugu rubrip	174	6	3.1	168	5	Q97194	Q97194 leishmania
102	6	3.1	84	16	Q99RK2	Q99rk2 staphylococ	175	6	3.1	168	5	Q903G9	Q903g9 caenorhabdi
103	6	3.1	90	16	Q9L381	Q9l381 rhizobium l	176	6	3.1	168	10	O96433	O96433 glycine max
104	6	3.1	91	16	Q98K21	Q98k21 rhizobium l	177	6	3.1	168	16	Q92110	Q92110 rickettsia
105	6	3.1	95	15	Q91C76	Q91ct6 chimpanzee	178	6	3.1	169	2	Q8VMP9	Q8vmp9 pseudomonas
106	6	3.1	96	5	Q95QF8	Q95qf8 caenorhabdi	179	6	3.1	170	10	O8W2E2	O8w2e2 glycine max
107	6	3.1	96	15	Q90CLO	Q90c10 human immun	180	6	3.1	170	10	Q8W2E1	Q8w2e1 glycine max
108	6	3.1	98	5	Q20560	Q20560 caenorhabdi	181	6	3.1	170	16	Q9JSH0	Q9jsh0 chlamydia p
109	6	3.1	102	16	O32370	O32370 campylobact	182	6	3.1	171	10	O9LZS2	O9lzs2 arabidopsis
110	6	3.1	104	17	Q96Y32	Q96y32 sulfolobus	183	6	3.1	171	10	O940P0	O940p0 arabidopsis
111	6	3.1	105	2	Q9FB68	Q9fb68 lactococcus	184	6	3.1	173	16	O9PJ08	O9pj08 campylobact
112	6	3.1	106	2	Q93GG6	Q93gg6 bacillus su	185	6	3.1	173	16	O9FJR6	O9fjr6 arabidopsis
113	6	3.1	107	16	Q9Z7V4	Q9z7v4 chlamydia p	186	6	3.1	174	10	O9FXJ1	O9fxj1 anabaena sp
114	6	3.1	107	16	Q93GB8	Q93gp8 salmonella	187	6	3.1	174	16	O97AM6	O97am6 thermoplasm
115	6	3.1	107	16	O52124	O52124 escherichia	188	6	3.1	175	17	O98LFO	O98lp0 rhizobium l
116	6	3.1	108	2	O52854	O52854 bacillus su	189	6	3.1	180	2	Q9APV2	Q9apv2 pseudomonas
117	6	3.1	108	5	O8SXF6	O8sxf6 drosophila	190	6	3.1	180	16	O8Y616	O8y616 listeria mo
118	6	3.1	109	16	O9PD16	Q9pd16 xylella fas	191	6	3.1	181	10	O9LZL2	O9lzl2 arabidopsis
119	6	3.1	111	2	O53895	Q53895 streptomyce	192	6	3.1	188	6	Q28946	Q28946 sus scrofa
120	6	3.1	111	4	Q96SB5	Q96sb5 homo sapien	193	6	3.1	188	16	Q8RAR5	Q8rar5 thermoanaer
121	6	3.1	111	16	O8XLB3	O8xlb3 clostridium	194	6	3.1	189	16	Q8R5R2	Q8r5r2 thermoanaer
122	6	3.1	112	2	O53512	Q53512 lactococcus	195	6	3.1	189	17	O980V6	O980v6 sulfolobus
123	6	3.1	114	16	Q9CBP0	Q9cbp0 mycobacteri	196	6	3.1	191	5	O95S83	O95s83 drosophila
124	6	3.1	116	2	O69105	O69105 bacillus sp	197	6	3.1	191	5	O9XTN9	O9xtn9 meloidogyne
125	6	3.1	117	17	Q973M2	Q973m2 sulfolobus	198	6	3.1	193	4	Q96JQ8	Q96jq8 homo sapien
126	6	3.1	118	5	Q904E5	Q904e5 glossina pa	199	6	3.1	194	2	O9L5B5	O9l5b5 streptococ
127	6	3.1	119	12	O41527	O41527 canine herp	200	6	3.1	196	2	O66087	O66087 lactococcc
128	6	3.1	120	16	O8XDH2	O8xdh2 escherichia	201	6	3.1	196	16	Q97R41	Q97r41 streptococ
129	6	3.1	122	12	Q9WD91	Q9wd91 influenzavi	202	6	3.1	197	5	O8T671	O8t671 echinococu
130	6	3.1	122	16	O8ZR44	O8zr44 salmonella	203	6	3.1	198	17	O59386	O59386 pyrococcus
131	6	3.1	122	16	O8Z8M4	O8z8m4 salmonella	204	6	3.1	201	2	Q45797	Q45797 bacteroides
132	6	3.1	124	5	O9BL63	Q9bl63 caenorhabdi	205	6	3.1	201	4	Q90UX1	Q90ux1 homo sapien
133	6	3.1	126	5	Q9V8F1	Q9v8f1 drosophila	206	6	3.1	202	11	Q8R012	Q8r012 mus musculu
134	6	3.1	126	13	Q92130	Q92130 xenopus lae	207	6	3.1	202	16	O8YNI7	O8yni7 anabaena sp
135	6	3.1	126	13	Q92131	Q92131 xenopus lae	208	6	3.1	203	16	O50828	O50828 borrelia bu
136	6	3.1	127	2	O9RQH8	Q9rqh8 listeria mo	209	6	3.1	204	5	O15971	O15971 drosophila
137	6	3.1	127	9	O9XIR1	Q9xir1 bacterioph	210	6	3.1	205	2	P71426	P71426 klebsiella
138	6	3.1	127	16	Q8Y771	Q8y771 listeria mo	211	6	3.1	205	2	O32716	O32716 klebsiella
139	6	3.1	132	11	Q8RIC2	Q8ric2 mus musculu	212	6	3.1	206	5	O9VQ06	O9vq06 drosophila
140	6	3.1	132	16	Q9CDX7	Q9cdx7 lactococcus	213	6	3.1	206	16	O88045	O88045 streptomyce
141	6	3.1	133	5	O8T6D8	Q8t6d8 echinococu	214	6	3.1	207	5	O45781	O45781 caenorhabdi
142	6	3.1	133	5	O8T667	Q8t667 echinococu	215	6	3.1	209	10	O88XT3	O8xrt3 arabidopsis
143	6	3.1	134	5	O95U06	Q95uq6 branchiost	216	6	3.1	209	16	Q98A43	Q98a43 rhizobium l
144	6	3.1	135	5	O77007	Q77007 plasmodium	217	6	3.1	211	17	Q97616	Q97616 sulfolobus
145	6	3.1	135	5	O9TY38	Q9ty38 plasmodium	218	6	3.1	211	17	Q97616	Q97616 sulfolobus
146	6	3.1	135	5	O9TY37	Q9ty37 plasmodium	219	6	3.1	214	2	O51895	O51895 prochloroco
147	6	3.1	138	2	Q93SV5	Q93sv5 chlorobium	220	6	3.1	215	13	Q9DFS6	Q9dfs6 brachydanio
148	6	3.1	141	17	Q8TZG6	Q8tzg6 pyrococcus	221	6	3.1	216	8	O8SL92	O8sl92 euglena gra
149	6	3.1	148	17	Q90ZM2	Q90zm2 pyrococcus	222	6	3.1	218	12	O9C7R9	O9c7r9 arabidopsis
150	6	3.1	150	17	Q972E0	Q972e0 sulfolobus	223	6	3.1	218	10	O55589	O55589 avian rotav
151	6	3.1	152	2	Q939G6	Q939g6 pseudomonas	224	6	3.1	219	5	O8TIE9	O8tie9 dictyosteli
152	6	3.1	153	2	O48287	Q48287 haemophilus	225	6	3.1	220	2	O9XCW1	O9xcw1 escherichia
153	6	3.1	153	5	O8WRQ4	Q8wrq4 solenopsis	226	6	3.1	221	11	O99NX6	O99nx6 tamias stri
154	6	3.1	153	5	O8WRQ3	Q8wrq3 solenopsis	227	6	3.1	221	11	O8VEJ3	O8vej3 mus musculu
155	6	3.1	153	5	O8WRQ2	Q8wrq2 solenopsis	228	6	3.1	221	16	Q97KD2	Q97kd2 clostridium
156	6	3.1	153	5	O8WRQ1	Q8wrq1 solenopsis	229	6	3.1	222	3	O01599	O01599 phanerocha
157	6	3.1	153	5	O8WRQ0	Q8wrq0 solenopsis	230	6	3.1	224	16	O9PJ95	O9pj95 campylobact
158	6	3.1	153	5	O8WRP9	Q8wrp9 solenopsis	231	6	3.1	225	3	O74194	O74194 candida par
159	6	3.1	153	5	O8WRP8	Q8wrp8 solenopsis	232	6	3.1	225	5	O967U1	O967u1 dictyosteli
160	6	3.1	153	5	O8WRP7	Q8wrp7 solenopsis	233	6	3.1	225	11	O99NX1	O99nx1 cricetus
161	6	3.1	153	5	O8WP92	Q8wp92 solenopsis	234	6	3.1	226	5	P91481	P91481 caenorhabdi
162	6	3.1	153	5	O8WP90	Q8wp90 solenopsis	235	6	3.1	226	10	O9FY40	O9fy40 ipomoea bat
										227	17	Q9YB38	Q9yb38 aeropyrum p

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2003, 09:57:38 : Search time 31 Seconds

(without alignments)
1276.163 Million cell updates/sec

Title: US-09-868-352-23

Perfect score: 192

Sequence: 1 MRKEVTPMLNKNKYPQPF.....EEQDKEMTSAKQHLLFVRKN 192

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

SPTREMBL_21.*

1: sp.archaea.*

2: sp.bacteria.*

3: sp.fungi.*

4: sp.human.*

5: sp.invertebrate.*

6: sp.mammal.*

7: sp.mhc.*

8: sp.organelle.*

9: sp.phage.*

10: sp.plant.*

11: sp.rodent.*

12: sp.virus.*

13: sp.vertebrate.*

14: sp.unclassified.*

15: sp.rvirus.*

16: sp.bacteriaph.*

17: sp.archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	12	6.2	102	16	O32127 bacillus su
2	12	6.2	170	16	Q97RN6 streptococ
3	12	6.2	214	16	Q99YU4 streptococ
4	11	5.7	91	16	Q928N1 listeria mo
5	11	5.7	93	16	Q9K7D1 bacillus ha
6	8	4.2	109	5	Q21089 caenorhabdi
7	8	4.2	251	16	Q9CJ28 lactococcus
8	8	4.2	502	3	Q12418 saccharomyc
9	8	4.2	733	4	Q96J38 homo sapien
10	7	3.6	60	2	Q93U42 carsonella
11	7	3.6	60	2	Q93U40 carsonella
12	7	3.6	88	12	Q9YNZ7 choristoneu
13	7	3.6	127	16	Q9SVF1 staphylococ
14	7	3.6	129	15	Q9QFQ7 human immun
15	7	3.6	137	5	Q8STE7 caenorhabdi
16	7	3.6	151	4	Q9H6C7 homo sapien

17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89

160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000

O67360 aquifex aeo
O9nze5 homo sapien
Q9396 streptococ
O23308 arabidopsis
Q9svp3 arabidopsis
Q8sw53 encephalito
Q9m003 arabidopsis
Q9n100 eptatretus
Q8vz95 arabidopsis
Q96qk2 homo sapien
Q9bi97 caenorhabdi
Q9chh2 thermoanaer
Q9chh6 mus musculu
Q9pvz0 xenopus lae
Q9xsc1 bos taurus
O14780 homo sapien
Q96fz7 homo sapien
Q96fn8 homo sapien
Q9dbp1 mus musculu
Q92lk3 mus musculu
Q9ed3 rhizobium l
Q8ujf4 agrobacteri
Q9zpi8 lycopersico
O12298 feline coro
Q88515 transmissib
Q9vra4 transmissib
Q9dy21 transmissib
Q9lw03 transmissib
Q9f7d6 uncultured
Q9zrb8 deinococcus
O51626 borrelia bu
O8xna1 clostridium
O01159 caenorhabdi
Q12066 saccharomyc
Q18652 caenorhabdi
O74232 gibberella
Q97ld8 sulfolobus
Q9a3z9 caulobacter
Q22040 caenorhabdi
Q9ciz7 lactococcus
Q975s8 sulfolobus
Q92ip3 rickettsia
Q9fpw0 chlamydomon
Q9zq47 arabidopsis
Q98q63 mycoplasma
O14587 homo sapien
Q91615 xenopus lae
O17232 caenorhabdi
O48547 zea mays (m
Q9gvr2 rattus sp.
O44088 stylonychia
Q9df49 xenopus lae
O44055 stylonychia
Q9hxm4 pseudomonas
Q8s648 oryza sativ
P79773 gallus gall
Q28224 cercopithec
Q9q96 soil-borne
Q8w5t9 bacterioph
Q96rr2 homo sapien
Q9y615 homo sapien
Q9bz0 homo sapien
Q96r94 homo sapien
Q96r95 homo sapien
Q90za0 gallus gall
O61139 cryptospori
Q9jad2 soil-borne
Q26032 plasmodium
Q8wtr6 homo sapien
Q9zm22 magnetospir
Q9b97 canis famill
Q922y6 mus musculu


```

RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kojonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA."
RL Nucleic Acids Res. 28:2311-2314(2000).
CC -!- FUNCTION: THIS PROTEIN IS ONE OF THE EARLY ASSEMBLY PROTEINS OF
CC THE 50S RIBOSOMAL SUBUNIT (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE L13P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001610; AAD18400.1; -.
DR EMBL; AE002211; AAF38342.1; -.
DR EMBL; AP002545; BAA98457.1; -.
DR TIGR; CP0515; -.
DR InterPro; IPR001074; Ribosomal_L13.
DR Pfam; PF00572; Ribosomal_L13; 1.
DR ProDom; PD001791; Ribosomal_L13; 1.
DR TIGRFAMs; TIGR01066; rplM_bact; 1.
DR PROSITE; PS00783; RIBOSOMAL_L13; FALSE_NEG.
KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 149 AA; 16940 MW; 0EFECF02FAA492DF CRC64;
Query Match 3.1%; Score 6; DB 1; Length 149;
Best Local Similarity 100.0%; Pred. NO. 59;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 159 KRKDTK 164
Db 3 KRKDTK 8

```

Search completed: February 26, 2003, 10:00:20
Job time : 37 secs

RESULT 39

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A. (GENE CLUSTER XIH3).
 RX MEDLINE=86037224; PubMed=3863963;
 RA Perry M., Thomsen G.H., Roeder R.G.;
 RT "Genomic organization and nucleotide sequence of two distinct histone
 gene clusters from *Xenopus laevis*. Identification of novel conserved
 upstream sequence elements.";
 RL J. Mol. Biol. 185:479-499(1985).
 RN [2]
 RP SEQUENCE OF 1-29 AND 62-84.
 RC TISSUE=Erythrocyte;
 RX MEDLINE=78144893; PubMed=638193;
 RA van Helden P., Strickland W.N., Brandt W.F., von Holt C.;
 RT "Histone H2B variants from the erythrocytes of an amphibian, a
 reptile and a bird.";
 RL Biochim. Biophys. Acta 533:278-281(1978).
 CC -!- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF
 H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATELY 146 BP OF DNA.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- SIMILARITY: BELONGS TO THE HISTONE H2B FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X03018; CAA26816.1; -;
 DR EMBL; M21287; AAA49768.1; -;
 DR PIR; A24510; HSXLB1.
 DR InterPro; IPR000558; Histone_H2B.
 DR Pfam; PF00125; histone; 1.
 DR PRINTS; PR00621; HISTONEH2B.
 DR PRODOM; PD000497; Histone_H2B; 1.
 DR SMART; SM00427; H2B; 1.
 DR PROSITE; PS00357; Histone_H2B; 1.
 DR Nucleic acid protein; Chromosomal protein; Nucleosome core; DNA-binding;
 KW Multigene family.
 FT INIT_MET 0
 FT SEQUENCE 125 AA; 13803 MW; DA6C1248381E6F94 CRC64;
 Query Match 3.1%; Score 6; DB 1; Length 125;
 Best Local Similarity 100.0%; Pred. No. 50;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 127 KRRKSR 132
 Db 28 KRRKSR 33
 RESULT 36
 ID H2B2_XENLA STANDARD; PRT; 125 AA.
 AC P06900;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Histone H2B.2.
 OS *Xenopus laevis* (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A. (GENE CLUSTER XIH1).
 RX MEDLINE=86037224; PubMed=3863963;

RA Perry M., Thomsen G.H., Roeder R.G.;
 RT "Genomic organization and nucleotide sequence of two distinct histone
 gene clusters from *Xenopus laevis*. Identification of novel conserved
 upstream sequence elements.";
 RL J. Mol. Biol. 185:479-499(1985).
 CC -!- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF
 H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATELY 146 BP OF DNA.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- SIMILARITY: BELONGS TO THE HISTONE H2B FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X03017; CAA26811.1; -;
 DR EMBL; M21286; AAA49763.1; -;
 DR PIR; B24510; HSXLB2.
 DR InterPro; IPR000558; Histone_H2B.
 DR Pfam; PF00125; histone; 1.
 DR PRINTS; PR00621; HISTONEH2B.
 DR PRODOM; PD000497; Histone_H2B; 1.
 DR SMART; SM00427; H2B; 1.
 DR PROSITE; PS00357; Histone_H2B; 1.
 DR Nucleic acid protein; Chromosomal protein; Nucleosome core; DNA-binding;
 KW Multigene family.
 FT INIT_MET 0
 FT SEQUENCE 125 AA; 13774 MW; 24A035A3F438CA95 CRC64;
 Query Match 3.1%; Score 6; DB 1; Length 125;
 Best Local Similarity 100.0%; Pred. No. 50;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 127 KRRKSR 132
 Db 28 KRRKSR 33
 RESULT 37
 ID GP48_BPSP1 STANDARD; PRT; 126 AA.
 AC O48402;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Putative gene 48 protein.
 GN 48.
 OS Bacteriophage SP01.
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
 OC SP01-like viruses.
 OX NCBI_TaxID=10685;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98327781; PubMed=9657951;
 RA Stewart C.R., Gaslightwala I., Hinata K., Krolkowski K.A.,
 RA Needleman D.S., Peng A.S.-Y., Peterman M.A., Tobias A., Wei P.;
 RT "Genes and regulatory sites of the 'host-takeover module' in the
 terminal redundancy of *Bacillus subtilis* bacteriophage SP01.";
 RL Virology 246:329-340(1998).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF031901; AAC29017.1; -;

RT the extremely thermophilic archaeon, Sulfolobus solfataricus P2. ";

RL Genome 43:116-136(2000).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 35092 / DSM 1617 / P2;

RX MEDLINE=21332296; PubMed=11427726;

RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,

RA Wayez M.J., Chau-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,

RA De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,

RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,

RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,

RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,

RA Garrett R.A., Ragan M.A., Sengen C.W., Van der Oost J.;

RA "The complete genome of the crenarchaeon Sulfolobus solfataricus P2. ";

RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).

CC -1- SIMILARITY: BELONGS TO THE UPF0148 FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: Y18930; CAB57522.1; -.

DR EMBL: AE006701; AA41078.1; -.

DR InterPro: IPR005355; UPF0148.

DR Pfam: PF03680; UPF0148; 1.

KW Hypothetical protein; Complete proteome.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 118 AA; 13445 MW; C92FE8819473D8B2 CRC64;

Query Match 3.1%; Score 6; DB 1; Length 118;

Best Local Similarity 100.0%; Pred. No. 48;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 26 IVKSDD 31

Db 52 IVKSDD 57

|||||

RESULT 33

ID YIBI_ECOLI STANDARD; PRT; 120 AA.

AC P32108;

DT 01-OCT-1993 (Rel. 27, Created)

DT 01-OCT-1993 (Rel. 27, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical protein yibi.

DE YIBI OR B3598.

GN Escherichia coli.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

OX NCBI_TaxID=562;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=K12;

RX MEDLINE=93259920; PubMed=8387990;

RA Zhao S., Sandt C.H., Feulner G., Vlazny D.A., Gray J.A., Hill C.W.;

RT "Rhs elements of Escherichia coli K-12: complex composites of shared

RT and unique components that have different evolutionary histories.;"

RN J. Bacteriol. 175:2799-2808(1993).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=K12 / MG1655;

RX MEDLINE=94316500; PubMed=8041620;

RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;

RT "Analysis of the Escherichia coli genome. V. DNA sequence of the

RT region from 76.0 to 81.5 minutes.;"

RL Nucleic Acids Res. 22:2576-2586(1994).

CC -1- SIMILARITY: STRONG, TO E.COLI YIAW.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: L19044; AAC95069.1; -.

DR EMBL: U00039; AAB18575.1; -.

DR EMBL: AE000437; AAC76622.1; -.

DR Ecogene; EG11765; yibi.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 120 AA; 13866 MW; 207110BBE2C03F67 CRC64;

Query Match 3.1%; Score 6; DB 1; Length 120;

Best Local Similarity 100.0%; Pred. No. 49;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 89 RISRL 94

Db 98 RISRL 103

|||||

RESULT 34

YLB9_CAEEL STANDARD; PRT; 122 AA.

AC P46583;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 01-NOV-1995 (Rel. 32, Last annotation update)

DE Hypothetical 13.8 kDa protein C34E10.9 in chromosome III.

GN C34E10.9

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;

RA Kirsten J.;

RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: U10402; AAA19071.1; -.

DR Wormpep; C34E10.9; CE01189.

KW Hypothetical protein.

SQ SEQUENCE 122 AA; 13766 MW; F30B8CD153D95598 CRC64;

Query Match 3.1%; Score 6; DB 1; Length 122;

Best Local Similarity 100.0%; Pred. No. 49;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 RQERH 170

Db 46 RQERH 51

|||||

RESULT 35

H2B1_XENLA STANDARD; PRT; 125 AA.

ID H2B1_XENLA

AC P02281;

DT 21-JUL-1986 (Rel. 01, Created)

DT 01-JAN-1988 (Rel. 06, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE Histone H2B.1.

OS Xenopus laevis (African clawed frog).

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: AE000796; AAB84540.1; -;
DR InterPro: IPR000302; KOW_motif.
DR Pfam: PF00467; KOW; 1.
KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 75 AA; 8335 MW; 4F84864F6F1258D CRC64;

Query Match 3.1%; Score 6; DB 1; Length 75;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 KSDIE 33
Db 60 KSDIE 65

RESULT 30

PHIL_MYTED
ID PHIL_MYTED STANDARD; PRT; 91 AA.
AC Q04621;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Sperm-specific protein PHI-1.

OS Mytilus edulis (Blue mussel).
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloida;
OC Mytiloidea; Mytilidae; Mytilus.
OX NCBI_TaxID=6550;

RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Sperm;
RX MEDLINE=93324352; PubMed=8332480;
RA Ruiz-Lara S., Prats E., Casas M.T., Cornudella L.;
RT "Molecular cloning and sequence of a cDNA for the sperm-specific
RT protein phi 1 from the mussel Mytilus edulis.";
RL Nucleic Acids Res. 21:2774-2774(1993).

CC -!- FUNCTION: INVOLVED IN NUCLEAR BASIC PROTEIN TRANSITION: HISTONES
CC ARE REPLACED BY SPERMATID SPECIFIC PROTEINS WHICH ARE THEMSELVES
CC REPLACED BY PROTAMINS IN LATE SPERMATIDS.
CC -!- SUBCELLULAR LOCATION: Nuclear.

CC -!- TISSUE SPECIFICITY: Sperm.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: X69718; CA449375.1; -;
DR PIR: S31482; S31482.
DR PIR: S34115; S34115.
KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
KW Nuclear protein.
FT INIT_MET 0

SQ SEQUENCE 91 AA; 10367 MW; FC7E6F5F28344626 CRC64;

Query Match 3.1%; Score 6; DB 1; Length 91;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 KRSRKS 135
Db 9 KRSRKS 14

RESULT 31

RT "Gene content and organization of a 281-kbp contig from the genome of

CVAY_RICCN
ID CVAY_RICCN STANDARD; PRT; 103 AA.
AC Q92IH7;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cyay protein.
GN CVAY OR RC0443.

OS Rickettsia conorii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsieae; Rickettsia.
OX NCBI_TaxID=781;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=Malish 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL Science 293:2093-2098(2001).
CC -!- SIMILARITY: BELONGS TO THE FRATAXIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: AE008608; AAL02981.1; -;
DR InterPro: IPR002908; Frataxin_like.
DR Pfam: PF01491; Frataxin_Cyay; 1.
DR ProDom: PD006646; Frataxin_like; 1.
DR PROSITE: PS01344; FRATAXIN_1; 1.
DR PROSITE: PS00810; FRATAXIN_2; 1.
KW Complete proteome.
SQ SEQUENCE 103 AA; 11689 MW; 1B3017B3E27FA40D CRC64;

Query Match 3.1%; Score 6; DB 1; Length 103;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 173 EQDKE 178
Db 22 EQDKE 27

RESULT 32

Y781_SULSO
ID Y781_SULSO STANDARD; PRT; 118 AA.
AC Q9UXG8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein SSO0781.

GN SSO0781 OR C40.017.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2287;

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=20165948; PubMed=10701121;
RA Charlebois R.L., Singh R.K., Chan-Weiher C.C.-Y., Allard G., Chow C.,
RA Confalonieri F., Curtis B., Duquet M., Erasmo G., Fahey D.,
RA Gaasterland T., Garrett R.A., Gordon P., Jeffries A.C., Koza C.,
RA Kushwaha N., Lafleur E., Medina N., Peng X., Penny S.L., She Q.,
RA St Jean A., van der Oost J., Young F., Zivanovic Y., Doolittle W.F.,
RA Ragan M.A., Sensen C.W.;
RT "Gene content and organization of a 281-kbp contig from the genome of

```

Query Match      3.6%; Score 7; DB 1; Length 2386;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 33 EFQLVIN 39
    |||||
Db 1299 EFQLVIN 1305

RESULT 27
PHIL_MYTCA
ID PHIL_MYTCA STANDARD; PRT; 35 AA.
AC P35422;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DE Sperm-specific protein PHI-1 (PL-III) (Sperm-specific protamine-like
DE protein) (Fragment).
OS Mytilus californianus (California mussel).
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloida;
OC Mytiloida; Mytilidae; Mytilus.
OX NCBI_TaxID=6549;
RN [1]
RP SEQUENCE.
RC TISSUE=Sperm; PubMed=7677995;
RX MEDLINE=9310699; Borrell I., Hunt D.F., Ausio J.;
RA "Sequence and characterization of a sperm-specific histone H1-like
RT protein of Mytilus californianus."
RL J. Biol. Chem. 268:185-194(1993).
CC -1- FUNCTION: INVOLVED IN NUCLEAR BASIC PROTEIN TRANSITION: HISTONES
CC ARE REPLACED BY SPERMATID SPECIFIC PROTEINS WHICH ARE THEMSELVES
CC REPLACED BY PROTAMINES IN LATE SPERMATIDS.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: SPERM.
DR PIR; B45316; B45316.
KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
KW Nuclear protein.
FT NON_TER 35 35
SQ SEQUENCE 35 AA; 3759 MW; 97459D27A76F4BD1 CRC64;

Query Match      3.1%; Score 6; DB 1; Length 35;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 130 KSRKS 135
    |||||
Db 8 KSRKS 13

RESULT 28
YM83_MYCTU
ID YM83_MYCTU STANDARD; PRT; 64 AA.
AC Q50682;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Very hypothetical protein RV2283.
GN RV2283 OR MT2341.1 OR MTCY339.27C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,

Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
"Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence."
Nature 393:537-544(1998).
(2)
SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
"Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains."
Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z77163; CAB00985.1; -;
DR EMBL; AE007077; -; NOT_ANNOTATED_CDS.
DR TIGR; MT2341.1; -;
DR TubercuList; RV2283; -;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 64 AA; 6557 MW; 9B7C5467EB8451AC CRC64;

Query Match      3.1%; Score 6; DB 1; Length 64;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 81 ASTRK 86
    |||||
Db 33 ASTRK 38

RESULT 29
R14E_METH
ID R14E_METH STANDARD; PRT; 75 AA.
AC O26139;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 50S ribosomal protein L14E.
GN RPL14E OR MTH31.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Delta H;
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
"Complete genome sequence of Methanobacterium thermoautotrophicum
RT delta: functional analysis and comparative genomics."
PL J. Bacteriol. 179:7135-7155(1997).
CC -1- SIMILARITY: BELONGS TO THE L14E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Wallis J.M.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PUTATIVE RNA HELICASE INVOLVED IN THE SECOND STEP OF RNA
CC SPlicing (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- DOMAIN: COMPOSED OF TWO SIMILAR DOMAINS.
CC -!- SIMILARITY: BELONGS TO THE HELICASE FAMILY. SKI2 SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AL110485; CAB60351.1; -.
DR WormPep; Y46G5A.4; CE21971.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR004179; Sec63.
DR Pfam: PF00270; DEAD; 2.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF02889; Sec63; 2.
DR SMART: SM00487; DEXDC; 2.
DR SMART: SM00490; HELICC; 1.
DR KW Hypothetical protein; Helicase; mRNA processing; mRNA splicing;
KW Spliceosome; Nuclear protein; ATP-binding; Repeat.
FT DOMAIN 497 897
FT NP_BIND 1344 1738
FT NP_BIND 497 504
FT NP_BIND 1344 1351
FT SITE 609 612
FT SITE 1448 1451
FT SITE 221 228
FT DOMAIN 2145 2438
FT SEQUENCE 2145 AA; 243829 MW; 2C1EE0F69583891F CRC64;
QY 36 LVINEKS 42
Db 2084 LVINEKS 2090
Query Match
Best Local Similarity 100.0%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 36 LVINEKS 42
Db 2084 LVINEKS 2090
RESULT 26
RAD3_SCHPO STANDARD; PRT: 2386 AA.
AC Q02099; Q9UUM1; Q92391;
DT 01-JUL-1993 (Rel. 26, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE DNA repair protein rad3.
GN RAD3 OR SPBC216.05.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN=972;
RX MEDLINE=97133293; PubMed=8978690;
RA Bentley N.J., Holtzman D.A., Flaggs G., Keegan K.S., DeMaggio A.,
RA Ford J.C., Hoekstra M., Carr A.M.;
RT "The Schizosaccharomyces pombe rad3 checkpoint gene.";
RL EMBO J. 15:6641-6651(1996).
RN [2]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
RA Welljans I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Frizic C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
RN [3]
RP SEQUENCE OF 711-1781 FROM N.A.
RX MEDLINE=93012978; PubMed=1398093;
RA Seaton B.L., Yucel J., Sunnerhagen P., Subramani S.;
RT "Isolation and characterization of the Schizosaccharomyces pombe rad3
RL gene, involved in the DNA damage and DNA synthesis checkpoints.";
CC Gene 119:83-89(1992).
CC -!- FUNCTION: INVOLVED IN G2 ARREST FOLLOWING DNA DAMAGE WHERE IT
CC PHOSPHORYLATES CHK1. IT IS ALSO INVOLVED IN THE DEPENDENCE OF
CC MITOSIS ON THE COMPLETION OF DNA REPLICATION.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE PI3/P14-KINASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Y09076; CAA70297.1; -.
DR EMBL; U76307; AAC49607.1; -.
DR EMBL; AL049558; CAB40165.1; -.
DR EMBL; X63544; CAA45106.1; -.
DR PIR; S25834; S25834.
DR InterPro: IPR003151; FAT.
DR InterPro: IPR003152; FATC.
DR InterPro: IPR000403; PI3_P14_kinase.
DR InterPro: IPR001440; TPR.
DR Pfam: PF00454; PI3_P14_kinase; 1.
DR Pfam: PF02259; FAT; 1.
DR Pfam: PF02260; FATC; 1.
DR SMART; SM00146; PI3KC; 1.
DR PROSITE; PS00915; PI3_4_KINASE_1; FALSE_NEG.
DR PROSITE; PS00916; PI3_4_KINASE_2; FALSE_NEG.
DR PROSITE; PS0290; PI3_4_KINASE_3; 1.
DR Transferrase; Kinase; DNA damage; DNA synthesis; DNA repair;
KW Nuclear protein.
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN=972;
RX MEDLINE=97133293; PubMed=8978690;
RA Bentley N.J., Holtzman D.A., Flaggs G., Keegan K.S., DeMaggio A.,
RA Ford J.C., Hoekstra M., Carr A.M.;
RT "The Schizosaccharomyces pombe rad3 checkpoint gene.";
RL EMBO J. 15:6641-6651(1996).
RN [2]

```

```

RA Glasheen E., Lane W.S., Pierce J.H., White M.F.;
RL "Role of IRS-2 in insulin and cytokine signalling.";
CC Nature 377:173-177(1995).
CC -1- FUNCTION: MAY MEDIATE THE CONTROL OF VARIOUS CELLULAR PROCESSES
CC BY INSULIN.
CC -1- TISSUE SPECIFICITY: SKELETAL MUSCLE, LUNG, BRAIN, LIVER, KIDNEY,
CC HEART AND SPLEEN.
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 PTB DOMAIN.
CC HSP; P35568; IIRS.
DR MGD; MGI:109334; IIRS2.
DR InterPro; IPR002404; Insulin_receptsl.
DR InterPro; IPR001849; PH.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF02174; IRS; 1.
DR PRINTS; PR00628; INSULINRSI.
DR SMART; SM00233; PH; 1.
DR SMART; SM00310; PTBI; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
KW Phosphorylation.
FT DOMAIN 16 144 PH.
FT DOMAIN 187 300 PTB.
FT MOD_RES 536 536 (BY SIMILARITY).
FT MOD_RES 649 649 (BY SIMILARITY).
FT MOD_RES 671 671 (BY SIMILARITY).
FT MOD_RES 911 911 (BY SIMILARITY).
FT MOD_RES 970 970 (BY SIMILARITY).
FT MOD_RES 1242 1242 (BY SIMILARITY).
FT MOD_RES 1303 1303 (BY SIMILARITY).
FT DOMAIN 19 28 POLY-ASN.
FT DOMAIN 444 449 POLY-SER.
FT DOMAIN 638 641 POLY-SER.
FT DOMAIN 936 939 POLY-SER.
SQ SEQUENCE 1321 AA; 136526 MW; 5069CE9D614960C7 CRC64;

Query Match 3.6%; Score 7; DB 1; Length 1321;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 RSKSQSS 138
Db 300 RSKSQSS 306

RESULT 24
IRS2_HUMAN STANDARD; PRT; 1324 AA.
AC Q9Y4H2;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Insulin receptor substrate-2 (IRS-2).
GN IRS2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97460123; PubMed=9312143;
RA Ogihara T., Isobe T., Ichimura T., Taoka M., Funaki M., Sakoda H.,
RA Onishi Y., Inukai K., Anai M., Fukushima Y., Kikuchi M., Yazaki Y.,
RA Oka Y., Asano T.;
RT "14-3-3 protein binds to insulin receptor substrate-1, one of the
RT binding sites of which is in the phosphorytyrosine binding domain.";
RL J. Biol. Chem. 272:25267-25274(1997).

```

```

CC -1- FUNCTION: MAY MEDIATE THE CONTROL OF VARIOUS CELLULAR PROCESSES
CC BY INSULIN.
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 PTB DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB000732; BAA24500.1; -.
DR HSSP; P35568; IIRS.
DR Genew; HGNC:6126; IRS2.
DR MIM; 600797; -.
DR InterPro; IPR002404; Insulin_receptsl.
DR InterPro; IPR001849; PH.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF02174; IRS; 1.
DR PRINTS; PR00628; INSULINRSI.
DR SMART; SM00233; PH; 1.
DR SMART; SM00310; PTBI; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
KW Phosphorylation.
FT DOMAIN 16 144 PH.
FT DOMAIN 190 303 PTB.
FT MOD_RES 540 540 (BY SIMILARITY).
FT MOD_RES 653 653 (BY SIMILARITY).
FT MOD_RES 675 675 (BY SIMILARITY).
FT MOD_RES 919 919 (BY SIMILARITY).
FT MOD_RES 978 978 (BY SIMILARITY).
FT MOD_RES 1253 1253 (BY SIMILARITY).
FT DOMAIN 19 28 POLY-ASN.
FT DOMAIN 371 380 POLY-ALA.
FT DOMAIN 447 452 POLY-SER.
FT DOMAIN 460 467 POLY-PRO.
FT DOMAIN 533 537 POLY-GLY.
FT DOMAIN 642 645 POLY-SER.
FT DOMAIN 694 701 POLY-ALA.
FT DOMAIN 944 947 POLY-SER.
FT DOMAIN 1031 1038 POLY-PRO.
FT DOMAIN 1265 1278 POLY-PRO.
SQ SEQUENCE 1324 AA; 136482 MW; 3D7B4AB2AE45104 CRC64;

Query Match 3.6%; Score 7; DB 1; Length 1324;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 RSKSQSS 138
Db 303 RSKSQSS 309

RESULT 25
U520_CAEEL STANDARD; PRT; 2145 AA.
ID U520_CAEEL
AC Q9U2G0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative U5 small nuclear ribonucleoprotein 200 kDa helicase.
GN Y46GSA.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.

```



```
CC -----S85963; AAB21608.1; -.
DR EMBL; S85963; AAB21608.1; -.
DR EMBL; S62539; AAB27175.1; -.
DR PIR; JS0670; JS0670.
DR PDB; 1JRS; 15-MAY-97.
DR Genew; HGNC:6125; IRS1.
DR MIM; 147545; -.
DR InterPro; IPR002404; Insulin_receptS1.
DR InterPro; IPR001849; PH.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF02174; IRS; 1.
DR PRINTS; PR00628; INSULINRS1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00310; PTBI; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
DR Phosphorylation; Polymorphism; Diabetes mellitus; Disease mutation;
KW 3D-structure.
FT DOMAIN 12 115 PH.
FT DOMAIN 157 267 PTB.
FT DOMAIN 128 134 POLY-GLY.
FT DOMAIN 391 398 POLY-SER.
FT DOMAIN 680 686 POLY-SER.
FT DOMAIN 807 815 POLY-SER.
FT DOMAIN 877 882 POLY-GLN.
FT DOMAIN 1055 1058 POLY-SER.
FT DOMAIN 1126 1130 POLY-GLY.
FT DOMAIN 1131 1135 POLY-SER.
FT DOMAIN 1197 1207 POLY-PRO.
FT MOD_RES 99 99 PHOSPHORYLATION (BY CK2)
FT MOD_RES 455 455 (BY SIMILARITY).
FT MOD_RES 612 612 PHOSPHORYLATION (BY INSR)
FT MOD_RES 632 632 PHOSPHORYLATION (BY INSR)
FT MOD_RES 896 896 PHOSPHORYLATION (BY INSR)
FT MOD_RES 941 941 PHOSPHORYLATION (BY INSR)
FT MOD_RES 989 989 PHOSPHORYLATION (BY INSR)
FT MOD_RES 1179 1179 PHOSPHORYLATION (BY INSR)
FT MOD_RES 1229 1229 PHOSPHORYLATION (BY INSR)
FT VARIANT 512 512 A -> P.
FT VARIANT 723 723 /FTIG=VAR_005299.
FT VARIANT 971 971 MISSING (IN NIDDM).
FT VARIANT 1043 1043 /FTIG=VAR_005301.
FT VARIANT 1095 1095 G -> R.
FT VARIANT 1095 1095 /FTIG=VAR_005300.
FT VARIANT 1095 1095 S -> Y (IN NIDDM).
FT VARIANT 1095 1095 /FTIG=VAR_005302.
FT VARIANT 1095 1095 C -> Y (IN NIDDM).
FT VARIANT 1095 1095 /FTIG=VAR_005303.
FT CONFLICT 134 134 G -> GG (IN REF. 2).
FT CONFLICT 362 362 S -> R (IN REF. 2).
FT CONFLICT 384 384 P -> R (IN REF. 2).
SQ SEQUENCE 1242 AA; 131590 MW; 3C0EFD9E32B3E64A CRC64;

Query Match 3.6%; Score 7; DB 1; Length 1242;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 RSKSQSS 138
Db 267 RSKSQSS 273
IIIIIII

RESULT 22
VIVD_BPT7 STANDARD; PRT; 1318 AA.
ID VIVD_BPT7
```

```
AC P03726;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Internal virion protein D.
GN 16.
OS Bacteriophage T7.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;
OC T7-like viruses.
OX NCBI_TaxID=10760;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83241725; PubMed=6864790;
RT Dunn J.J., Studier F.W.;
RT "Complete nucleotide sequence of bacteriophage T7 DNA and the
RT locations of T7 genetic elements.";
RL J. Mol. Biol. 166:477-535(1983).
RN [2]
RP SIMILARITY TO SLT.
RX MEDLINE=94262160; PubMed=8203016;
RA Koonin E.V., Rudd K.E.;
RT "A conserved domain in putative bacterial and bacteriophage
RT transglycosylases.";
RL Trends Biochem. Sci. 19:106-107(1994).
CC -!- FUNCTION: MAY BE INVOLVED IN THE LYSIS OF THE BACTERIAL CELL
CC WALL DURING THE RELEASE OF THE PHAGE PROGENY.
CC -!- SIMILARITY: BELONGS TO THE SLT FAMILY OF TRANSGLYCOSYLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by 'and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; V01146; CAA24434.1; -.
DR PIR; A04352; HIBPD7.
DR PIR; S42332; S42332.
DR InterPro; IPR000189; SLT_domain.
DR Pfam; PF01464; SLT; 1.
DR PROSITE; PS00922; TRANSGLYCOSYLASE; 1.
KW Cell wall; Hydrolase; Glycosidase.
FT DOMAIN 24 111 SLT-TYPE DOMAIN.
FT ACT_SITE 37 37 BY SIMILARITY.
SQ SEQUENCE 1318 AA; 143838 MW; 51A0AAA920CBF210 CRC64;

Query Match 3.6%; Score 7; DB 1; Length 1318;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 KEVTPEM 9
Db 720 KEVTPEM 726
IIIIIII

RESULT 23
IRS2_MOUSE STANDARD; PRT; 1321 AA.
ID IRS2_MOUSE
AC P81122;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Insulin receptor substrate-2 (IRS-2) (4PS).
GN IRS2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=95405472; PubMed=7675087;
RA Sun X.J., Wang L.-M., Zhang Y., Yenush L., Myers M.G. Jr.,
```

```

RT cloning.";
RL Biochim. Biophys. Acta 1172:323-326(1993).
CC -!- FUNCTION: MAY MEDIATE THE CONTROL OF VARIOUS CELLULAR PROCESSES
CC BY INSULIN. WHEN PHOSPHORYLATED BY THE INSULIN RECEPTOR BINDS
CC SPECIFICALLY TO VARIOUS CELLULAR PROTEINS CONTAINING SH2 DOMAINS
CC SUCH AS PHOSPHATIDYLINOSITOL 3-KINASE P85 SUBUNIT OR GRB-2.
CC -!- SIMILARITY: CONTAINS 1 PTB DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PTB DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L24563; AAA9335.1; -.
DR EMBL; X69722; CAA49378.1; -.
DR PIR; S43514; S43514.
DR HSP; P35568; IIRS.
DR MGD; MGI:99454; Irs1.
DR InterPro; IPR002404; Insulin_receptS1.
DR InterPro; IPR001849; PH.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF02174; IRS; 1.
DR PRINTS; PR00628; INSULINRS1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00310; PTBI; 1.
DR PROSITE; PSS0003; PH_DOMAIN; 1.
KW Phosphorylation.
FT DOMAIN 12 115 PH.
FT DOMAIN 152 262 PTB.
FT DOMAIN 675 680 POLY-SER.
FT DOMAIN 872 877 POLY-GLN.
FT DOMAIN 1119 1128 POLY-GLY.
FT DOMAIN 1194 1198 POLY-PRO.
FT MOD_RES 99 99 PHOSPHORYLATION (BY CK2)
FT MOD_RES 460 460 PHOSPHORYLATION (BY INSR)
FT MOD_RES 608 608 PHOSPHORYLATION (BY INSR)
FT MOD_RES 628 628 PHOSPHORYLATION (BY INSR)
FT MOD_RES 891 891 PHOSPHORYLATION (BY INSR)
FT MOD_RES 935 935 PHOSPHORYLATION (BY INSR)
FT MOD_RES 983 983 PHOSPHORYLATION (BY INSR)
FT MOD_RES 1173 1173 PHOSPHORYLATION (BY INSR)
FT MOD_RES 1220 1220 PHOSPHORYLATION (BY INSR)
FT MOD_RES 1038 1039 PHOSPHORYLATION (BY INSR)
FT MOD_RES 1182 1182 PHOSPHORYLATION (BY INSR)
FT MOD_RES 1233 AA; 130723 MW; C0E9B2D890DADD87 CRC64;
SQ SEQUENCE 1233 AA; 130723 MW; C0E9B2D890DADD87 CRC64;

Query Match 3.6% Score 7; DB 1; Length 1233;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 RSKSQSS 138
Db 262 RSKSQSS 268
|||||||

RESULT 21
IRSI_HUMAN
ID IRS1_HUMAN STANDARD; PRT; 1242 AA.
AC P35568;
DT 01-JUN-1994 (Rel. 29, Created)

```

```

DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Insulin receptor substrate-1 (IRS-1).
GN IRS1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=93292738; PubMed=8513971;
RA Araki E., Sun X.J., Haag B.L. III, Chuang L.M., Zhang Y.,
RA Yang-Feng T.L., White M.F., Kahn C.R.;
RT "Human skeletal muscle insulin receptor substrate-1. Characterization
RT of the cDNA, gene, and chromosomal localization.";
RL Diabetes 42:1041-1054(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92181456; PubMed=1311924;
RA Nishiyama M., Wands J.R.;
RT "Cloning and increased expression of an insulin receptor substrate-1-
RT like gene in human hepatocellular carcinoma.";
RL Biochem. Biophys. Res. Commun. 183:280-285(1992).
RN [3]
RP VARIANTS PRO-512 AND ARG-971.
RX MEDLINE=93390176; PubMed=8104271;
RA Almind K., Bjoerbaek C., Vestergaard H., Hansen T., Echwald S.,
RA Pedersen O.;
RT "Aminoacid polymorphisms of insulin receptor substrate-1 in
RT non-insulin-dependent diabetes mellitus.";
RL Lancet 342:828-832(1993).
RN [4]
RP VARIANT NIDDM GLY-723 DEL.
RX MEDLINE=96303710; PubMed=8723689;
RA Esposito D.L., Mammarella S., Ranieri A., della Loggia F., Capani F.,
RA Consoli A., Mariani-Costantini R., Caramia F.G., Cama A., Battista P.;
RT "Deletion of Gly723 in the insulin receptor substrate-1 of a patient
RT with noninsulin-dependent diabetes mellitus.";
RL Hum. Mutat. 7:364-366(1996).
RN [5]
RP VARIANTS NIDDM TYR-1043 AND TYR-1095.
RA Mammarella S., Creati B., Esposito D.L., Arcuri P., della Loggia F.,
RA Capani F., Mariani-Costantini R., Caramia F.G., Battista P., Cama A.;
RT "Novel allele of the insulin receptor substrate-1 bearing two
RT non-conservative amino acid substitutions in a patient with
RT noninsulin-dependent diabetes mellitus.";
RL Hum. Mutat. 11:411-411(1998).
RN [6]
RP STRUCTURE BY NMR OF 157-267.
RX MEDLINE=96185451; PubMed=8599766;
RA Zhou M.-W., Huang B., Olejniczak E.T., Meadows R.P., Shuker S.B.,
RA Miyazaki M., Trueb T., Shoelson S.E., Fesik S.W.;
RT "Structural basis for IL-4 receptor phosphopeptide recognition by the
RT IRS-1 PTB domain.";
RL Nat. Struct. Biol. 3:388-393(1996).
CC -!- FUNCTION: MAY MEDIATE THE CONTROL OF VARIOUS CELLULAR PROCESSES
CC BY INSULIN. WHEN PHOSPHORYLATED BY THE INSULIN RECEPTOR BINDS
CC SPECIFICALLY TO VARIOUS CELLULAR PROTEINS CONTAINING SH2 DOMAINS
CC SUCH AS PHOSPHATIDYLINOSITOL 3-KINASE P85 SUBUNIT OR GRB-2.
CC -!- DISEASE: POLYMORPHISMS IN IRS1 MAY BE INVOLVED IN THE ETIOLOGY OF
CC A SUBSET OF LATE-ONSET NON-INSULIN-DEPENDENT DIABETES MELLITUS
CC (NIDDM).
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PTB DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

FT DOMAIN 272 321 PHORBOL-ESTER AND DAG BINDING 1.
FT DOMAIN 337 383 PHORBOL-ESTER AND DAG BINDING 2.
FT DOMAIN 432 558 CATALYTIC-A (POTENTIAL).
FT DOMAIN 578 752 CATALYTIC-B (POTENTIAL).
FT VARSPLOC 451 475 MISSING (IN SHORT ISOFORM).
SQ SEQUENCE 791 AA; 88996 MW; C7DD07F5B285FF62 CRC64;

Query Match 3.6%; Score 7; DB 1; Length 791;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 RRSRSK 134
Db 784 RRSRSK 790

RESULT 18
L100_ADEGX STANDARD; PRT; 798 AA.
AC F36856;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Late 100 kDa protein.
OS Avian adenovirus gal10 (strain SA2) (Fowl adenovirus 10).
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Aviadenovirus.
OX NCBI_TaxID=10547;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94040780; PubMed=8224879;
RA Sheppard M.;
RT "Identification of a fowl adenovirus gene with sequence homology to the 100K gene of human adenovirus.";
RL Gene 132:307-308(1993).
CC -!- FUNCTION: THE 100 kDa PROTEIN IS A LATE NONSTRUCTURAL PROTEIN INVOLVED IN TRANSPORT OF HEXON FROM CYTOPLASM TO THE NUCLEUS (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L07890; AAA72328.1; -
CC PIR; JN0878;
CC InterPro; IPR003381; Adeno_100.
CC Pfam; PF02438; adeno_100; 1.
KW Late protein; Transport.
SQ SEQUENCE 798 AA; 89028 MW; 04CFA042AC6A0DF4 CRC64;

Query Match 3.6%; Score 7; DB 1; Length 798;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 ENPNRP 117
Db 592 ENPNRP 598

RESULT 19
SYL_BUCAI STANDARD; PRT; 859 AA.
ID P57519;
AC P57519;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (Leurs).
GN LRSU OR BU444.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum symbiotic bacterium).

```

```

OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Tokyo 1998;
MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. APS.";
RL Nature 407:81-86(2000).
CC -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP + diphosphate + L-leucyl-tRNA(Leu).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AP001119; BAB13142.1; -
CC InterPro; IPR002302; Leu-TRNASyntla.
CC InterPro; IPR002300; tRNA-synt_1a.
CC InterPro; IPR001412; tRNA-synt_1.
CC Pfam; PF00133; tRNA-synt_1; 1.
CC PRINTS; PR00985; TRNASYNTHLEU.
CC TIGRfams; TIGR00396; leuS_bact; 1.
CC PROSITE; PS00178; AA-TRNA_LIGASE_1; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding; Complete proteome.
FT SITE 42 52 "HIGH" REGION.
FT SITE 618 622 "KMSKS" REGION.
FT BINDING 621 621 ATP (BY SIMILARITY).
SQ SEQUENCE 859 AA; 102170 MW; 43DB88CD4020975 CRC64;

Query Match 3.6%; Score 7; DB 1; Length 859;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 IKEEQDK 177
Db 741 IKEEQDK 747

RESULT 20
IRSI_MOUSE STANDARD; PRT; 1233 AA.
ID IRSI_MOUSE
AC P35569;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Insulin receptor substrate-1.
GN IRSI OR IRS-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94220494; PubMed=81671159;
RA Araki E., Haag B.L., III, Kahn C.R.;
RT "Cloning of the mouse insulin receptor substrate-1 (IRS-1) gene and complete sequence of mouse IRS-1.";
RL Biochim. Biophys. Acta 1221:353-356(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93192326; PubMed=8448209;
RA Keller S.R., Abersold R., Garner C.W., Lienhard G.E.;
RT "The insulin-elicited 160 kDa phosphotyrosine protein in mouse adipocytes is an insulin receptor substrate 1: identification by

```

DR Pfam: PF00130; DAG_PE-bind; 2.
DR Pfam: PF00609; DAGKa; 1.
DR Pfam: PF00781; DAGKc; 1.
DR PRINTS: PR00008; DAGPEDOMAIN.
DR ProDom: PD000012; EF-hand; 1.
DR ProDom: PD002939; DAGKa; 1.
DR ProDom: PD005043; DAG_kin_cat; 1.
DR SMART: SM00109; C1; 2.
DR SMART: SM00045; DAGKa; 1.
DR SMART: SM00046; DAGKc; 1.
DR SMART: SM00054; EFh; 2.
DR PROSITE: PS00479; DAG_PE_BIND_DOM_1; 2.
DR PROSITE: PS00081; DAG_PE_BIND_DOM_2; 2.
DR PROSITE: PS00018; EF_HAND; 1.
DR Transferase: Kinase; Calcium-binding; Phorbol-ester binding;
KW Repeat; Multigene family.
FT CA_BIND 185 196 EF-HAND 1 (POTENTIAL).
FT CA_BIND 230 241 EF-HAND 2 (POTENTIAL).
FT DOMAIN 269 318 PHORBOL-ESTER AND DAG BINDING 1.
FT DOMAIN 334 380 PHORBOL-ESTER AND DAG BINDING 2.
FT DOMAIN 429 555 CATALYTIC-A (POTENTIAL).
FT DOMAIN 575 749 CATALYTIC-B (POTENTIAL).
SQ SEQUENCE 788 AA; 88521 MW; E77BCA40B22CID49 CRC64;
Query Match 3.6%; Score 7; DB 1; Length 788;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 128 RRSRSK 134
| | | | |
Db 781 RRSRSK 787
RESULT 17
KDGG_HUMAN STANDARD; PRT; 791 AA.
AC P49619;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Diacylglycerol kinase, gamma (EC 2.7.1.107) (Diglyceride kinase) (DGK-
DE gamma) (DAG kinase gamma).
GN DGK OR DAGK3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=94308084; PubMed=8034597;
RA Kai M., Sakane F., Imai S.-I., Wada I., Kanoh H.;
RT "Molecular cloning of a diacylglycerol kinase isozyme predominantly
RT expressed in human retina with a truncated and inactive enzyme
RT expression in most other human cells.";
RL J. Biol. Chem. 269:18492-18498(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99168758; PubMed=10071200;
RA Stoeckl H., Klein J., Gehrig A., Koehler M.R., Jurkles B., Kellner U.,
RA Leo-Kottler B., Schmid M., Weber B.H.F.;
RT "Mapping and genomic characterization of the gene encoding
RT diacylglycerol kinase gamma (DAGK3): assessment of its role in
RT dominant optic atrophy (OPAL).";
RL Hum. Genet. 104:99-105(1999).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=99168758; PubMed=10071200;
RA Stoeckl H., Klein J., Gehrig A., Koehler M.R., Jurkles B., Kellner U.,
RA Leo-Kottler B., Schmid M., Weber B.H.F.;
RT "Mapping and genomic characterization of the gene encoding
RT diacylglycerol kinase gamma (DAGK3): assessment of its role in
RT dominant optic atrophy (OPAL).";
RL Hum. Genet. 104:99-105(1999).
CC -!- FUNCTION: REVERSES THE NORMAL FLOW OF GLYCEROLIPID
CC BIOSYNTHESIS BY PHOSPHORYLATING DIACYLGLYCEROL BACK TO
CC PHOSPHATIDIC ACID.
CC -!- CATALYTIC ACTIVITY: ATP + 1,2-diacylglycerol = ADP + 1,2-
CC diacylglycerol 3-phosphate.
CC -!- ENZYME REGULATION: REQUIRES PHOSPHATIDYL SERINE FOR MAXIMAL
CC ACTIVITY.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC. CAN BE LOOSELY BOUND TO THE

MEMBRANES.
-!- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
SHORT FORM: ARE PRODUCED BY ALTERNATIVE SPLICING. THE SHORT FORM
MAY BE INACTIVE.
-!- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN RETINA AND IN A
MUCH LESSER EXTENT IN THE BRAIN. OTHER TISSUES CONTAIN EXTREMELY
LOW LEVELS OF DGK-GAMMA.
-!- SIMILARITY: BELONGS TO THE EUKARYOTIC DIACYLGLYCEROL KINASE
FAMILY.
-!- SIMILARITY: CONTAINS 2 ZINC-DEPENDENT PHORBOL-ESTER AND DAG
BINDING DOMAINS.
-!- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: D26135; BAA05132.1; -
DR EMBL: AF020945; AAC04686.1; -
DR EMBL: AF020922; AAC04686.1; JOINED.
DR EMBL: AF020923; AAC04686.1; JOINED.
DR EMBL: AF020924; AAC04686.1; JOINED.
DR EMBL: AF020925; AAC04686.1; JOINED.
DR EMBL: AF020926; AAC04686.1; JOINED.
DR EMBL: AF020927; AAC04686.1; JOINED.
DR EMBL: AF020928; AAC04686.1; JOINED.
DR EMBL: AF020929; AAC04686.1; JOINED.
DR EMBL: AF020930; AAC04686.1; JOINED.
DR EMBL: AF020931; AAC04686.1; JOINED.
DR EMBL: AF020932; AAC04686.1; JOINED.
DR EMBL: AF020933; AAC04686.1; JOINED.
DR EMBL: AF020934; AAC04686.1; JOINED.
DR EMBL: AF020935; AAC04686.1; JOINED.
DR EMBL: AF020936; AAC04686.1; JOINED.
DR EMBL: AF020937; AAC04686.1; JOINED.
DR EMBL: AF020938; AAC04686.1; JOINED.
DR EMBL: AF020939; AAC04686.1; JOINED.
DR EMBL: AF020940; AAC04686.1; JOINED.
DR EMBL: AF020941; AAC04686.1; JOINED.
DR EMBL: AF020942; AAC04686.1; JOINED.
DR EMBL: AF020943; AAC04686.1; JOINED.
DR EMBL: AF020944; AAC04686.1; JOINED.
DR Genew: HGNC:2853; DGKc.
DR MIM: 601854; -
DR InterPro: IPR000756; DAGKa.
DR InterPro: IPR001206; DAGKc.
DR InterPro: IPR002219; DAG_PE-bind.
DR InterPro: IPR003622; DAG_kin_cat.
DR InterPro: IPR002046; EF-hand.
DR Pfam: PF00036; efhand; 2.
DR Pfam: PF00130; DAG_PE-bind; 2.
DR Pfam: PF00609; DAGKa; 1.
DR Pfam: PF00781; DAGKc; 1.
DR PRINTS: PR00008; DAGPEDOMAIN.
DR ProDom: PD000012; EF-hand; 1.
DR ProDom: PD002939; DAGKa; 1.
DR ProDom: PD005043; DAG_kin_cat; 1.
DR SMART: SM00109; C1; 2.
DR SMART: SM00045; DAGKa; 1.
DR SMART: SM00046; DAGKc; 1.
DR SMART: SM00054; EFh; 2.
DR PROSITE: PS00479; DAG_PE_BIND_DOM_1; 2.
DR PROSITE: PS00081; DAG_PE_BIND_DOM_2; 2.
DR PROSITE: PS00018; EF_HAND; 2.
KW Transferase: Kinase; Calcium-binding; Phorbol-ester binding;
KW Repeat; Multigene family; Alternative splicing.
FT DOMAIN 151 156 POLY-SER.
FT CA_BIND 188 199 EF-HAND 1 (POTENTIAL).
FT CA_BIND 233 244 EF-HAND 2 (POTENTIAL).

```
DR PROSITE; PS00372; PTS_EI1A_2; 1.
KW Phosphotransferase system; Sugar transport; Transferase;
KW Transmembrane; Phosphorylation; Complete proteome.
FT DOMAIN 1 333 EIIC DOMAIN.
FT DOMAIN 334 484 EIIB DOMAIN.
FT DOMAIN 485 631 EIIA DOMAIN.
FT TRANSMEM 20 40 POTENTIAL.
FT TRANSMEM 52 72 POTENTIAL.
FT TRANSMEM 86 106 POTENTIAL.
FT TRANSMEM 134 154 POTENTIAL.
FT TRANSMEM 161 181 POTENTIAL.
FT TRANSMEM 213 233 POTENTIAL.
FT TRANSMEM 269 289 POTENTIAL.
FT TRANSMEM 313 333 POTENTIAL.
FT MOD_RES 256 256 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 380 380 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 548 548 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 632 AA; 69974 MW; DC413636A0BFF52 CRC64;

Query Match 3.6%; Score 7; DB 1; Length 632;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 ILLKYDF 62
      |||||
Db 332 ILLKYDF 338

RESULT 15
KGG_MOUSE
ID KGG_MOUSE STANDARD; PRT; 788 AA.
AC Q91WG7;
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Diacylglycerol kinase, gamma (EC 2.7.1.107) (Diglyceride kinase) (DGK-
DE gamma) (DAG kinase gamma) (88 kDa diacylglycerol kinase).
GN DGK OR DGK3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Kidney;
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: ATP + 1,2-diacylglycerol -> ADP + 1,2-
CC diacylglycerol 3-phosphate.
CC -1- SIMILARITY: BELONGS TO THE EUKARYOTIC DIACYLGLYCEROL KINASE
CC FAMILY.
CC -1- SIMILARITY: CONTAINS 2 ZINC-DEPENDENT PHORBOL-ESTER AND DAG
CC BINDING DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; BC015278; AAH15278.1; -.
CC MGD; MGI:105060; Dagk3.
CC InterPro; IPR000756; DAGK.
CC InterPro; IPR001206; DAGKc.
CC InterPro; IPR002219; DAG_PE-bind.
CC InterPro; IPR003622; DAG_kin_cat.
CC InterPro; IPR002048; EF-hand.
CC Pfam; PF00036; ehand; 2.
CC Pfam; PF00130; DAG_PE-bind; 2.
CC Pfam; PF00609; DAGKa; 1.

DR PROSITE; PS00781; DAGKc; 1.
DR ProDom; PD000012; EF-hand; 1.
DR ProDom; PD002939; DAGKa; 1.
DR ProDom; PD005043; DAG_kin_cat; 1.
DR PROSITE; PS00479; DAG_PE_BIND_DOM_1; 2.
DR PROSITE; PS00081; DAG_PE_BIND_DOM_2; 2.
DR PROSITE; PS00018; EF_HAND; 2.
KW Transferase; Kinase; Calcium-binding; Phorbol-ester binding;
KW Repeat; Multigene family.
FT CA_BIND 185 196 EF-HAND 1 (POTENTIAL).
FT CA_BIND 230 241 EF-HAND 2 (POTENTIAL).
FT DOMAIN 269 318 PHORBOL-ESTER AND DAG BINDING 1.
FT DOMAIN 334 380 PHORBOL-ESTER AND DAG BINDING 2.
FT DOMAIN 429 555 CATALYTIC-A (POTENTIAL).
FT DOMAIN 575 749 CATALYTIC-B (POTENTIAL).
SQ SEQUENCE 788 AA; 88523 MW; 969298DFD144F5DF CRC64;

Query Match 3.6%; Score 7; DB 1; Length 788;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 RRKRSRK 134
      |||||
Db 781 RRKRSRK 787

RESULT 16
KGG_RAT
ID KGG_RAT STANDARD; PRT; 788 AA.
AC P49620;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Diacylglycerol kinase, gamma (EC 2.7.1.107) (Diglyceride kinase) (DGK-
DE gamma) (DAG kinase gamma) (88 kDa diacylglycerol kinase).
GN DGK OR DGK3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar;
RX MEDLINE=95108095; PubMed=7809169;
RA Goto K., Funayama M., Kondo H.;
RA "Cloning and expression of a cytoskeleton-associated diacylglycerol
RA kinase that is dominantly expressed in cerebellum.";
RL Proc Natl Acad Sci U S A. 91:13042-13046(1994).
CC -1- CATALYTIC ACTIVITY: ATP + 1,2-diacylglycerol -> ADP + 1,2-
CC diacylglycerol 3-phosphate.
CC -1- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN BRAIN AND LOCALIZED
CC PREDOMINANTLY IN CEREBELLAR PURKINJE CELLS.
CC -1- SIMILARITY: BELONGS TO THE EUKARYOTIC DIACYLGLYCEROL KINASE
CC FAMILY.
CC -1- SIMILARITY: CONTAINS 2 ZINC-DEPENDENT PHORBOL-ESTER AND DAG
CC BINDING DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D38448; BAA07480.1; -.
CC InterPro; IPR000756; DAGKa.
CC InterPro; IPR001206; DAGKc.
CC InterPro; IPR002219; DAG_PE-bind.
CC InterPro; IPR003622; DAG_kin_cat.
CC InterPro; IPR002048; EF-hand.
CC Pfam; PF00036; ehand; 2.
CC Pfam; PF00036; ehand; 2.
```

```
Best Local Similarity 100.0%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 RKSRKS 135
DB 342 RKSRKS 348

RESULT 12
NCAP_CVPRM STANDARD; PRT; 382 AA.
AC P24411;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Nucleocapsid protein (N structural protein).
GN N.
OS Porcine respiratory coronavirus (strain RM4) (PRCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Coronaviridae; Coronavirus.
OX NCBI_TaxID=11148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91073120; PubMed=2174956;
RA Rasschaert D., Duarte M., Laude H.;
RT "Porcine respiratory coronavirus differs from transmissible
RT gastroenteritis virus by a few genomic deletions.";
RL J. Gen. Virol. 71:2599-2607(1990).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; Z24675; CAA80841.1; -.
DR PIR; E36607; E36607.
DR InterPro; IPR001218; Corona_nucleocap.
DR Pfam; PF00937; Corona_nucleoca; 1.
KW Nucleocapsid.
SQ SEQUENCE 382 AA; 43551 MW; 76EE7091E5D93BC9 CRC64;

Query Match 3.6%; Score 7; DB 1; Length 382;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 RKSRKS 135
DB 342 RKSRKS 348

RESULT 13
YX8_YEAST STANDARD; PRT; 577 AA.
AC P53075;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 66.9 kDa protein in SAP4-OST5 intergenic region.
GN YGL228W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Fartmann B., Kramer W.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO YEAST YPR039C.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
-----
```

```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; Z27250; CAA96945.1; -.
DR SGD; S0003197; SHE10.
KW Hypothetical protein.
SQ SEQUENCE 577 AA; 68862 MW; EF4CA95FA3157660 CRC64;

Query Match 3.6%; Score 7; DB 1; Length 577;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 KYDFIVG 65
DB 175 KYDFIVG 181

RESULT 14
PTMA_BUCAI STANDARD; PRT; 632 AA.
AC P57635;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE PTS system, mannitol-specific IIAB component (EIIABC-Mtl) (Mannitol-
DE permease IIAB component) (Phosphotransferase enzyme II, ABC
DE component) (EC 2.7.1.69) (EII-Mtl).
GN MTIA OR BU572.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS symbiotic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Tokyo 1998;
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
RT Buchnera sp. APS.";
RL Nature 407:81-86(2000).
CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE
CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC THE SUGAR (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein
CC histidine + sugar phosphate.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC -!- PTM: AN INTRAMOLECULAR PHOSPHOTRANSFER TAKES PLACES BETWEEN
CC HIS-548 AND CYS-380 (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS 1 PTS EIIA DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PTS EIIB DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PTS EIIC DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AP001119; BAB13262.1; -.
DR HSSP; P00550; IA3A.
DR InterPro; IPR002178; PTS_EIIA_2.
DR Pfam; PF00359; PTS_EIIA_2; 1.
DR ProDom; PD001689; PTS_EIIA_2; 1.
```

DT 01-NOV-1988 (Rel. 09, Last sequence update)
 DT 01-APR-1990 (Rel. 14, Last annotation update)
 DE Nucleocapsid protein (N structural protein).
 GN N.
 OS Porcine transmissible gastroenteritis coronavirus (strain FS772/70).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
 OC Coronaviridae; Coronavirus.
 OX NCBI_TaxID=11150;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88216185; PubMed=2835592;
 RA Britton P., Carmenes R.S., Page K.W., Garwes D.J., Parra F.;
 RT "Sequence of the nucleoprotein gene from a virulent British field
 isolate of transmissible gastroenteritis virus and its expression in
 RT Saccharomyces cerevisiae.";
 RL Mol. Microbiol. 2:89-99(1988).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Y00542; CAA68607.1; -.
 DR PIR; S03762; S03762.
 DR InterPro; IPR001218; Corona_nucleocap.
 DR Pfam; PF00937; Corona_nucleoca; 1.
 KW Nucleocapsid.
 SQ SEQUENCE 382 AA; 43483 MW; 25152E1D8F327F9E CRC64;
 Query Match 3.6%; Score 7; DB 1; Length 382;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 129 RKSRSKS 135
 Db 342 RKSRSKS 348
 |||||
 RESULT 10
 NCAP_CVPPU STANDARD; PRT; 382 AA.
 ID NCAP_CVPPU
 AC P04134;
 DT 01-NOV-1986 (Rel. 03, Created)
 DT 01-NOV-1988 (Rel. 09, Last sequence update)
 DT 01-APR-1990 (Rel. 14, Last annotation update)
 DE Nucleocapsid protein (N structural protein).
 GN N.
 OS Porcine transmissible gastroenteritis coronavirus (strain Purdue).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
 OC Coronaviridae; Coronavirus.
 OX NCBI_TaxID=11151;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88078100; PubMed=2825819;
 RA Rasschaert D., Gelfi J., Laude H.;
 RT "Enteric coronavirus TGEV: partial sequence of the genomic RNA, its
 RT organization and expression.";
 RL Biochimie 69:591-600(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86181608; PubMed=3008432;
 RA Kapke P.A., Brian D.A.;
 RT "Sequence analysis of the porcine transmissible gastroenteritis
 RT coronavirus nucleocapsid protein gene.";
 RL Virology 151:41-49(1986).
 RN [3]
 RP SEQUENCE OF 1-21 FROM N.A.
 RX MEDLINE=88306229; PubMed=2841792;
 RA Kapke P.A., Tung F.Y.T., Hogue B.G., Brian D.A., Woods R.D.,
 RA Wesley R.;

RT "The amino-terminal signal peptide on the porcine transmissible
 RT gastroenteritis coronavirus matrix protein is not an absolute
 DE requirement for membrane translocation and glycosylation.";
 GN Virology 165:367-376(1988).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X06371; CAA29674.1; -.
 DR EMBL; M14878; AAA47915.1; -.
 DR EMBL; M21627; AAA47913.1; -.
 DR PIR; A04025; VHIHPC.
 DR PIR; S03936; S03936.
 DR InterPro; IPR001218; Corona_nucleocap.
 DR Pfam; PF00937; Corona_nucleoca; 1.
 KW Nucleocapsid.
 FT CONFLICT 237 237 W -> S (IN REF. 2).
 FT CONFLICT 376 376 N -> I (IN REF. 2).
 SQ SEQUENCE 382 AA; 43522 MW; E299502A0FB36EAA CRC64;
 Query Match 3.6%; Score 7; DB 1; Length 382;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 129 RKSRSKS 135
 Db 342 RKSRSKS 348
 |||||
 RESULT 11
 NCAP_CVPR8 STANDARD; PRT; 382 AA.
 ID NCAP_CVPR8
 AC P33463;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-FEB-1994 (Rel. 28, Last annotation update)
 DE Nucleocapsid protein (N structural protein).
 GN N.
 OS Porcine respiratory coronavirus (strain 86/137004 / British isolate)
 OS (PRCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
 OC Coronaviridae; Coronavirus.
 OX NCBI_TaxID=33736;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92116634; PubMed=1662846;
 RA Britton P., Mawditt K.L., Page K.W.;
 RT "The cloning and sequencing of the virion protein genes from a
 RT British isolate of porcine respiratory coronavirus: comparison with
 RT transmissible gastroenteritis virus genes.";
 RL Virus Res. 21:181-198(1991).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X60056; CAA42657.1; -.
 DR PIR; S24282; S24282.
 DR InterPro; IPR001218; Corona_nucleocap.
 DR Pfam; PF00937; Corona_nucleoca; 1.
 KW Nucleocapsid.
 SQ SEQUENCE 382 AA; 43553 MW; 50350CED86FFCD4 CRC64;
 Query Match 3.6%; Score 7; DB 1; Length 382;

RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C.N., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gillwill R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Levaslatko M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConachie L.J., McKay K., McMurray A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann R.M., Symamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RT "The DNA sequence and comparative analysis of human chromosome 20.";
RL Nature 414:865-871(2001).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM SRP55-1).
RC TISSUE=Placenta;
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 21-27 AND 47-55.
RX MEDLINE=92249775; PubMed=1577277;
RA Zahler A.M., Lane W.S., Stolk J.A., Roth M.B.;
RT "SR proteins: a conserved family of pre-mRNA splicing factors.";
RL Genes Dev. 6:837-847(1992).
CC -1- FUNCTION: PLAYS A ROLE IN CONSTITUTIVE SPLICING AND CAN MODULATE
CC THE SELECTION OF ALTERNATIVE SPLICE SITES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS: SRP55-1 (SHOWN HERE), SRP55-
CC 2 AND SRP55-3; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- PTM: EXTENSIVELY PHOSPHORYLATED ON SERINE RESIDUES IN THE RS
CC DOMAIN (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).
CC -1- SIMILARITY: BELONGS TO THE SR FAMILY OF SPLICING FACTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U30883; AAA93073.1; -;
DR EMBL: U30828; AAA93071.1; -;
DR EMBL: U30829; AAA93072.1; -;
DR EMBL: AL031681; CAB43960.1; -;
DR EMBL: BC006832; AAH06832.1; -;
DR Genew; HGNC:10788; SFRS6.
DR MIM; 601944; -;
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 2.
DR SMART; SM00360; rrm; 2.
DR PROSITE; PS50102; RRM; 2.
DR PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
KW Nuclear protein; RNA-binding; mRNA splicing; Alternative splicing;
KW Repeat; Phosphorylation.
FT DOMAIN 1 72 RNA-BINDING (RRM) 1.
FT DOMAIN 87 90 GLY-RICH (HINGE REGION).
FT DOMAIN 110 183 RNA-BINDING (RRM) 2.
FT DOMAIN 184 343 ARG/SER-RICH (RS DOMAIN).
FT VARSPPLIC 86 135 SGGGGYSRRKTSGRDYGPPVRYEYRLIVLENLSSRSWODL
FT KDFMRQAGE -> MTNGAEAVSTEAKMTAFDPWPFLHLC

FT DPCPMTLWLTLPEAMTTAAFC (IN ISOFORM SRP55-
FT 2).
FT MISSING (IN ISOFORM SRP55-2).
FT RVSPPPKRATSRSRSRSRSRSSSRD -> LKLC A
FT RFMSQQTESLYSLASSC (IN ISOFORM SRP55-3).
FT R -> H (IN REF. 1).
SQ SEQUENCE 344 AA: 39586 MW: 72305506CE948B94 CRC64;
Query Match 3.6%; Score 7; DB 1; Length 344;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 129 RKSRKS 135
DB 241 RKSRKS 247
|||||||
RESULT 8
DHPL_BACSH STANDARD; PRT; 381 AA.
AC P23307;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Phenylalanine dehydrogenase (EC 1.4.1.20) (PheDH).
GN PDH.
OS Bacillus sphaericus.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1421;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88255879; PubMed=2838396;
RA Okazaki N., Hibino Y., Asano Y., Ohmori M., Numao N., Kondo K.;
RT "Cloning and nucleotide sequencing of phenylalanine dehydrogenase
RL gene of Bacillus sphaericus.";
RL Gene 63:337-341(1988).
CC -1- CATALYTIC ACTIVITY: L-phenylalanine + H(2)O + NAD(+) =
CC phenylpyruvate + NH(3) + NADH.
CC -1- SIMILARITY: BELONGS TO THE GLU/LEU/PHE/VAL DEHYDROGENASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M26661; AAA22646.1; -;
DR InterPro; IPR001625; GLFV_Dh.
DR Pfam; PF00208; GLFV_dehydrog; 1.
DR Pfam; PF02812; GLFV_dehydrog_N; 1.
DR PRINTS; PR00082; GLFV_DHGRNASE.
DR PROSITE; PS00074; GLFV_DEHYDROGENASE; 1.
DR KX Oxidoreductase; NAD.
FT ACT_SITE 91 BY SIMILARITY.
FT NP_BIND 191 197 NAD (POTENTIAL).
SQ SEQUENCE 381 AA: 41578 MW: D32450CE397ACA67 CRC64; ;
Query Match 3.6%; Score 7; DB 1; Length 381;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 26 IVKSDDI 32
DB 236 IVKSDDI 242
|||||||
RESULT 9
NCAP_CVPFS
ID NCAP_CVPFS STANDARD; PRT; 382 AA.
AC P05991;
DT 01-NOV-1988 (Rel. 09, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Sec-independent protein translocase protein tatB.
 GN TATB OR H10187.1.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Rd / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitchugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
 RT Rd.";
 RL Science 269:496-512(1995).
 CC -!- FUNCTION: REQUIRED FOR CORRECT LOCALIZATION OF PRECURSOR PROTEINS
 CC BEARING SIGNAL PEPTIDES WITH THE TWIN ARGININE CONSERVED MOTIF
 CC S/T-R-X-F-L-K. THIS SEC-INDEPENDENT PATHWAY IS TERMED TAT FOR
 CC TWIN-ARGININE TRANSLLOCATION SYSTEM. THIS SYSTEM MAINLY TRANSPORTS
 CC PROTEINS WITH BOUND COFACTORS THAT REQUIRE FOLDING PRIOR TO EXPORT
 CC (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: INNER-MEMBRANE BOUND (PROBABLE).
 CC -!- SIMILARITY: BELONGS TO THE TATB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U32704; -; NOT_ANNOTATED_CDS.
 DR TIGR: H10187.1; -;
 DR InterPro: IPR003998; TatB.
 DR PRINTS; PRO1506; TATBPROTEIN.
 DR TIGRFAMS; TIGR01014; 2a6401s01; 1.
 DR Transport; Protein transport; Translocation; Transmembrane;
 KW Inner membrane; Complete proteome.
 FT TRANSMEM 1 21 POTENTIAL.
 SQ SEQUENCE 186 AA; 20430 MW; 7FD2166488CA057F CRC64;
 Query Match 3.6%; Score 7; DB 1; Length 186;
 Best Local Similarity 100.0%; Pred. No. 6.3;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 133 SKSQSSK 139
 Db 177 SKSQSSK 183
 RESULT 6
 RACH_DICDI STANDARD; PRT; 200 AA.
 ID RACH_DICDI
 AC Q9GPR7;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE RAS-related protein rach.
 GN RACH.
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostellium.
 OX NCBI_TaxID=44689;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=AX4;
 RX MEDLINE=21127961; PubMed=11222756;
 RA Rivero F., Dislich H., Glockner G., Noegel A.A.;
 RT "The Dictyostelium discoideum family of Rho-related proteins.";
 RL Nucleic Acids Res. 29:1068-1079(2001).
 CC -!- SIMILARITY: BELONGS TO THE SMALL GTPASE SUPERFAMILY. RHO FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF310894; AAG45133.1; -;
 DR HSSP; P21181; IAM4.
 DR DictyDB; DD?????; rach.
 DR InterPro: IPR003578; GTPase_Rho.
 DR InterPro: IPR001230; Prenyl_site.
 DR InterPro: IPR001806; Ras_trnsmg.
 DR InterPro: IPR005225; Small_GTP.
 DR Pfam; PF00071; ras; 1.
 DR PRINTS; PR00449; RASTRNSFRMG.
 DR SMART; SM00174; RHO; 1.
 DR TIGRFAMS; TIGR00231; small_GTP; 1.
 KW GTP-binding; Prenylation; Lipoprotein.
 FT NP_BIND 11 18 GTP (BY SIMILARITY).
 FT NP_BIND 58 62 GTP (BY SIMILARITY).
 FT NP_BIND 117 120 GTP (BY SIMILARITY).
 FT DOMAIN 33 41 EFFECTOR REGION (POTENTIAL).
 FT LIPID 197 197 GERANYL-GERANYL (BY SIMILARITY).
 SQ SEQUENCE 200 AA; 22189 MW; 0852F4358EC6A648 CRC64;
 Query Match 3.6%; Score 7; DB 1; Length 200;
 Best Local Similarity 100.0%; Pred. No. 6.7;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3 KEVTPEM 9
 Db 130 KEVTPEM 136
 RESULT 7
 SFR6_HUMAN STANDARD; PRT; 344 AA.
 ID SFR6_HUMAN
 AC Q13247; Q13244; Q13245; Q9UJB8; Q96J06;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Splicing factor, arginine/serine-rich 6 (pre-mRNA splicing factor
 DE SRP55).
 GN SFRS6 OR SRP55.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RC TISSUE-Colon;
 RX MEDLINE=96016206; PubMed=7556075;
 RA Screaton G.R., Caceres J.F., Mayeda A., Bell M.V., Plebanski M.,
 RA Jackson D.G., Bell J.I., Krainer A.R.;
 RT "Identification and characterization of three members of the human SR
 RT family of pre-mRNA splicing factors.";
 RL EMBO J. 14:4336-4349(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=21638749; PubMed=11780052;
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
 RA Bailey J., Barlow K.F., Bates K.N., Beare D.M.,

```
RP SEQUENCE FROM N.A.
RC STRAIN-CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Deboy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Unayama L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE L32P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AL021999; -; NOT ANNOTATED CDS.
DR EMBL: AE006985; AAK45255.1; ALT_INIT.
DR TIGR: MT1007; -.
DR TubercuList; Rv0979.lc; -.
DR InterPro: IPR002677; Ribosomal_L32p.
DR Pfam: PF01783; Ribosomal_L32p; 1.
KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 57 AA; 6507 MW; 2EF45101DDBF5921 CRC64;

Query Match 3.6%; Score 7; DB 1; Length 57;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 127 KRRKRSRS 133
Db 5 KRRKRSRS 11

RESULT 3
SFR6_RABIT STANDARD; PRT; 81 AA.
ID SFR6_RABIT
AC O18776;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Splicing factor, arginine/serine-rich 6 (Pre-mRNA splicing factor
DE SRP55) (Fragment).
DE SFR56 OR SRP55.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-New Zealand white;
RA Brunet A., Henion G., Duranthon V., Renard J.P.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PLAYS A ROLE IN CONSTITUTIVE SPLICING AND CAN MODULATE
CC THE SELECTION OF ALTERNATIVE SPLICE SITES (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- PTM: EXTENSIVELY PHOSPHORYLATED ON SERINE RESIDUES IN THE RS
CC DOMAIN (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).
CC -!- SIMILARITY: BELONGS TO THE SR FAMILY OF SPLICING FACTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```

```
DR EMBL: AF011564; AAB66467.1; -.
DR InterPro: IPR000504; RNA_rec_mot.
DR PROSITE: PS50102; RRM; PARTIAL.
DR PROSITE: PS00030; RRM_RNP_1; PARTIAL.
KW Nuclear protein; RNA-binding; mRNA splicing; Repeat; Phosphorylation.
FT NON_TER 1 >81 ARG/SER-RICH (RS DOMAIN).
FT DOMAIN <1 >81
FT NON_TER 81
SQ SEQUENCE 81 AA; 9520 MW; 20D9F8A9D08C7CA0 CRC64;

Query Match 3.6%; Score 7; DB 1; Length 81;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 129 RKSRKRS 135
Db 39 RKSRKRS 45

RESULT 4
RR12_PORPU STANDARD; PRT; 124 AA.
ID RR12_PORPU
AC P51289;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Chloroplast 30S ribosomal protein S12.
GN RP512.
OS Porphyra purpurea.
OC Chloroplast.
OC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
OX NCBI_TaxID=2787;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Avonport;
RA Reith M.E., Munnholland J.;
RT "Complete nucleotide sequence of the Porphyra purpurea chloroplast
RT genome.";
RL Plant Mol. Biol. Rep. 13:333-335(1995).
CC -!- SIMILARITY: BELONGS TO THE S12P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U38804; AAC08175.1; -.
DR InterPro: IPR000230; Ribosomal_S12.
DR Pfam: PF00164; Ribosomal_S12; 1.
DR PRINTS: PR01034; RIBOSOMALS12.
DR PRODOM: PD000576; Ribosomal_S12; 1.
DR TIGRFAMs: TIGR00981; rpsL_bact; 1.
DR PROSITE: PS00055; RIBOSOMAL_S12; 1.
KW Ribosomal protein; Chloroplast.
SQ SEQUENCE 124 AA; 13904 MW; 6632BE623C11FB1 CRC64;

Query Match 3.6%; Score 7; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 128 RKSRSK 134
Db 110 RKSRSK 116

RESULT 5
TATB_HAETIN STANDARD; PRT; 186 AA.
ID TATB_HAETIN
AC P57047; P44559;
DT 01-NOV-1995 (Rel. 32, Created)
```

983 5 2.6 280 1 Y705_CHLPN Q927k0 chlamydia p
 984 5 2.6 280 1 YB00_YEAST P32788 saccharomyc
 985 5 2.6 281 1 CLCX_CAEEL Q9ngj7 caenorhabdi
 986 5 2.6 281 1 MSP1_PLAFN P13820 plasmodium
 987 5 2.6 281 1 NPS2_MOUSE O55126 mus musculu
 988 5 2.6 281 1 TNF6_HUMAN P48023 homo sapien
 989 5 2.6 282 1 ARG6_THEMA Q922a4 thermotoga
 990 5 2.6 282 1 LPXC_CHLPN Q927q2 chlamydia p
 991 5 2.6 282 1 TR2A_HUMAN Q13595 homo sapien
 992 5 2.6 283 1 COBD_PYRHO O58114 pyrococcus
 993 5 2.6 283 1 FNKC_ECOLI P14190 escherichia
 994 5 2.6 283 1 IPFL_HUMAN P52945 h insulin p
 995 5 2.6 283 1 IPFL_MESAU P70118 mesocricetu
 996 5 2.6 283 1 IPFL_RAT P52947 rattus norv
 997 5 2.6 283 1 MTD_METKA P94951 methanopyru
 998 5 2.6 283 1 SFRA_MOUSE Q60701 mus musculu
 999 5 2.6 283 1 YBAS_BACSU P55190 bacillus su
 1000 5 2.6 284 1 IPFL_MOUSE P52946 mus musculu

ALIGNMENTS

RESULT 1
 KG62_HUMAN
 ID KG62_HUMAN STANDARD; PRT; 733 AA.
 AC Q15349; Q15419; Q9UJN5;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Ribosomal protein S6 kinase alpha 2 (EC 2.7.1.-) (S6K-alpha 2) (90
 DE kda ribosomal protein S6 kinase alpha 2 (p90-RSK 2) (Ribosomal S6 kinase 3)
 DE (RSK-3) (pp90RSK3).
 GN RPS6KA2 OR RSK3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95349602; PubMed=7623830;
 RA Zhao Y., Bjoerbaek C., Weremowicz S., Morton C.C., Moller D.E.;
 RT "RSK3 encodes a novel pp90rsk isoform with a unique N-terminal
 RT sequence: growth factor-stimulated kinase function and nuclear
 RT translocation.";
 RL Mol. Cell. Biol. 15:4353-4363(1995).
 RN [2]
 RP SEQUENCE OF 1-540 FROM N.A.
 RX MEDLINE=94189676; PubMed=8141249;
 RA Moller D.E., Xia C.-H., Tang W., Zhu A.X., Jakubowski M.;
 RT "Human rsk isoforms: cloning and characterization of tissue-specific
 RT expression.";
 RL Am. J. Physiol. 266:C351-C359(1994).
 RN [3]
 RP SEQUENCE OF 101-733 FROM N.A.
 RA Palmer S.;
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: PHOSPHORYLATES A WIDE RANGE OF SUBSTRATES INCLUDING
 CC RIBOSOMAL PROTEIN S6. IMPLICATED IN THE ACTIVATION OF THE MITOGEN-
 CC ACTIVATED KINASE CASCADE.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN MANY TISSUES. HIGHEST EXPRESSION
 CC IN LUNG AND SKELETAL MUSCLE.
 CC -!- PTM: AUTOPHOSPHORYLATED.
 CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC S6 KINASE SUBFAMILY. CONTAINS TWO KINASE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X85106; CAA59427.1; -
 DR EMBL; L07598; AAC82496.1; ALT_INIT.
 DR EMBL; AL022069; CAB58228.1; -
 DR HSSP; Q63450; 1A06.
 DR Genew; HGNC:10431; RPS6KA2.
 DR MIM; 601685; -
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR000961; Pkinase-C.
 DR InterPro; IPR002290; Ser_thr_pkinase.
 DR Pfam; PF00069; pkinase; 2.
 DR Pfam; PF00433; pkinase_C; 1.
 DR Prodom; PD000001; Euk_pkinase; 2.
 DR SMART; SM00133; S_TK_X; 1.
 DR SMART; SM00220; S_TK; 2.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 2.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST; 2.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 2.
 KW Transferase; Serine/threonine-protein kinase; ATP-binding;
 KW Repeat; Multigene family; Phosphorylation; Nuclear protein.
 FT DOMAIN 59 318
 FT PROTEIN KINASE 1.
 FT DOMAIN 415 672
 FT PROTEIN KINASE 2.
 FT NP_BIND 65 73
 FT BINDING 91 91
 FT ACT_SITE 184 184
 FT NP_BIND 421 429
 FT BINDING 444 444
 FT ACT_SITE 532 532
 FT CONFLICT 256 256
 FT CONFLICT 269 269
 FT CONFLICT 339 339
 FT CONFLICT 447 447
 FT CONFLICT D -> G (IN REF. 2).
 SQ SEQUENCE 733 AA; 83252 MW; FC82095707F95033 CRC64;
 Query Match 4.2%; Score 8; DB 1; Length 733;
 Best Local Similarity 100.0%; Pred. No. 1.9;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 128 RRKRSKS 135
 Db 19 RRKRSKS 26
 |||||
 RESULT 2
 RL32_MYCTU STANDARD; PRT; 57 AA.
 ID RL32_MYCTU
 AC P58287;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE 50S ribosomal protein L32.
 GN RPMF OR RV0979.1C OR MT1007 OR MTV044.07c.1.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37Rv;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544(1998).
 RN [2]

837	5	2.6	251	1	PROB_CAMJE	Q9p129	campylobact	910	5	2.6	264	1	YLPK_CAEEL	P91301	caenorhabdi
838	5	2.6	251	1	RSOA_YEAST	P32905	saccharomyc	911	5	2.6	264	1	YNV6_YEAST	P53853	saccharomyc
839	5	2.6	251	1	RSOB_YEAST	P46654	saccharomyc	912	5	2.6	265	1	EFID_XENLA	P29693	xenopus lae
840	5	2.6	251	1	YQRS_CAEEL	Q09308	caenorhabdi	913	5	2.6	265	1	IOD2_HUMAN	Q92813	homo sapien
841	5	2.6	252	1	YQRS_CAEEL	Q26540	schistosoma	914	5	2.6	265	1	PPNK_BACHD	Q9K904	baicillus ha
842	5	2.6	252	1	PMI_CANAL	P31353	candida alb	915	5	2.6	265	1	RFAP_SALTU	Q06995	salmonella
843	5	2.6	252	1	UNG_HSVSA	Q01019	herpesvirus	916	5	2.6	265	1	RP1A_ARATH	Q92u38	arabidopsis
844	5	2.6	253	1	PHOT_LISIN	Q92dm1	listeria in	917	5	2.6	265	1	UCRI_SOLTU	P37841	solanan tub
845	5	2.6	253	1	PHOT_LISMO	P58724	listeria mo	918	5	2.6	265	1	Y280_MYCGE	P47522	mycoplasma
846	5	2.6	253	1	RSO_CANAL	Q42817	candida alb	919	5	2.6	265	1	YG09_STRPN	Q97PK0	streptococc
847	5	2.6	253	1	RS3_AMBME	P79891	ambystoma m	920	5	2.6	266	1	AROE_HELPJ	Q92jx8	helicobacte
848	5	2.6	253	1	SOJ_BACHD	Q9K5n0	bacillus ha	921	5	2.6	267	1	DPM1_YEAST	P14020	saccharomyc
849	5	2.6	253	1	TRPA_LACLA	Q01997	lactococcus	922	5	2.6	267	1	PROB_STRTR	P96488	streptococc
850	5	2.6	253	1	UPPS_CHLTR	O84456	chlamydia t	923	5	2.6	268	1	IL2A_CANEA	O62802	canis famil
851	5	2.6	253	1	YSO2_ACTAM	P29086	acidianus a	924	5	2.6	268	1	NO30_MEDTR	P93329	medicago tr
852	5	2.6	254	1	IDNO_ECOLI	P39345	escherichia	925	5	2.6	268	1	UCR3_TOBAC	P51133	nicotiana t
853	5	2.6	254	1	YC43_PORPU	P51264	porphyra pu	926	5	2.6	268	1	UCR5_TOBAC	P51135	nicotiana t
854	5	2.6	254	1	YQCI_BACSU	P45944	bacillus su	927	5	2.6	269	1	CYSQ_ACTAC	P70714	actinobacil
855	5	2.6	255	1	DLX1_MOUSE	Q64317	mus musculus	928	5	2.6	269	1	SFR5_RAT	Q09167	rattus norv
856	5	2.6	255	1	SOLA_YEAST	P53315	saccharomyc	929	5	2.6	269	1	SRPB_MOUSE	P47758	mus musculus
857	5	2.6	256	1	CFI_ELAIM	O65333	elaeagnus u	930	5	2.6	269	1	SUR4_FUGRU	O57590	fugu rubrip
858	5	2.6	256	1	FBP2_DROME	P54398	drosophila	931	5	2.6	269	1	SUR4_HUMAN	O15360	homo sapien
859	5	2.6	256	1	M3R_DICDI	P11872	dictyosteli	932	5	2.6	270	1	HRF5_MOUSE	Q64310	mus musculus
860	5	2.6	256	1	OMAT_HUMAN	Q16633	homo sapien	933	5	2.6	270	1	RSP4_DROME	P38979	drosophila
861	5	2.6	256	1	VFAT_HRSVA	P03419	human respi	934	5	2.6	270	1	SFR5_MOUSE	Q35326	mus musculus
862	5	2.6	256	1	YREC_SYN2	P19737	synchococc	935	5	2.6	270	1	ALYS_BPRIT	Q01036	herpesvirus
863	5	2.6	257	1	AAPR_RHILV	O52815	rhizobium l	936	5	2.6	270	1	WFGC_SCHPO	Q38135	bacterioph
864	5	2.6	257	1	FAEJ_ECOLI	P33784	escherichia	937	5	2.6	270	1	ES1_BRARE	Q9P7C5	schizosacch
865	5	2.6	257	1	LIVG_METJA	Q58863	methanococc	938	5	2.6	270	1	HRF5_MOUSE	Q90257	brachydanio
866	5	2.6	257	1	YBE3_YEAST	P38195	saccharomyc	939	5	2.6	270	1	RSP4_DROME	Q9CPX5	mus musculus
867	5	2.6	257	1	YL48_MYCLE	Q9CCE2	mycobacteri	940	5	2.6	270	1	SUR4_DROME	P38979	drosophila
868	5	2.6	258	1	DTC_MOUSE	P01881	mus musculus	941	5	2.6	270	1	Y041_CHLTR	O18405	drosophila
869	5	2.6	258	1	LPXA_PSEAE	Q9X594	pseudomonas	942	5	2.6	270	1	YD1J_SCHPO	O84045	chlamydia t
870	5	2.6	258	1	UCRI_TOBAC	P47929	nicotiana t	943	5	2.6	270	1	YI19_PSEAE	Q10247	schizosacch
871	5	2.6	258	1	VSP2_AGRKH	P47797	agkistrodon	944	5	2.6	271	1	HM8_XENLA	Q9HY16	pseudomonas
872	5	2.6	259	1	GUB_BACBR	P37073	bacillus br	945	5	2.6	271	1	KKAI_ECOLI	P14837	xenopus lae
873	5	2.6	259	1	PEBI_CAMJE	P45678	campylobact	946	5	2.6	271	1	KKAI_SALTU	P00551	escherichia
874	5	2.6	259	1	PSTB_PASMU	Q9Cn17	pasteurella	947	5	2.6	271	1	MIND_THEMEA	Q03447	salmonella
875	5	2.6	259	1	YPRK_BACHD	Q9K9u0	bacillus ha	948	5	2.6	271	1	NIE1_HUMAN	Q9X213	thermotoga
876	5	2.6	259	1	YPRD_ECOLI	P45753	escherichia	949	5	2.6	271	1	SPED_CLOPE	Q12981	homo sapien
877	5	2.6	260	1	COX3_RAT	P05505	rattus norv	950	5	2.6	271	1	SRPB_HUMAN	Q8Xmz0	clostridium
878	5	2.6	260	1	COX3_XENLA	P00419	xenopus lae	951	5	2.6	271	1	YHE9_YEAST	Q9Y5m8	homo sapien
879	5	2.6	260	1	DH88_MOUSE	Q00171	mus musculus	952	5	2.6	272	1	UCR2_TOBAC	P38722	saccharomyc
880	5	2.6	260	1	STCE_EMENI	Q00674	emericeila	953	5	2.6	272	1	MP74_YEAST	Q13243	homo sapien
881	5	2.6	260	1	VSP1_AGKAC	Q918x2	agkistrodon	954	5	2.6	273	1	PROB_STRPY	P51132	nicotiana t
882	5	2.6	260	1	VSPB_TRIGA	O13061	trimeresuru	955	5	2.6	273	1	TC1A_CAEER	P39015	saccharomyc
883	5	2.6	261	1	COX3_HUMAN	P00414	homo sapien	956	5	2.6	273	1	COAT_RBDV	Q99J17	streptococc
884	5	2.6	261	1	COX3_HYLLA	Q95707	hylobates l	957	5	2.6	273	1	UCRI_MAIZE	P35072	caenorhabdi
885	5	2.6	261	1	COX3_MOUSE	P00416	mus musculus	958	5	2.6	274	1	COAT_RBDV	P49727	zea mays
886	5	2.6	261	1	COX3_PETMA	Q35539	petromyzon	959	5	2.6	274	1	COX3_ALLMA	P23629	raspberry b
887	5	2.6	261	1	COX3_PONPA	P92696	pongo pygma	960	5	2.6	275	1	EP12_STAEP	P80439	allomyces m
888	5	2.6	261	1	VE39_NPVOP	Q05121	orgyia pseu	961	5	2.6	275	1	PH4H_RHILO	P30193	staphylococ
889	5	2.6	262	1	FBOL_BOMMO	P21828	bombyx mori	962	5	2.6	275	1	VA16_VACCV	Q98472	rhizobium l
890	5	2.6	262	1	FRGI_DROME	Q9Vwa8	drosophila	963	5	2.6	275	1	YRB2_SYN7P	P16710	vaccinia vi
891	5	2.6	262	1	IOD2_MOUSE	Q9Z1y9	mus musculus	964	5	2.6	276	1	PTP3_CHLEU	P46205	synchococc
892	5	2.6	262	1	IOD2_RAT	P70551	rattus norv	965	5	2.6	277	1	PYRF_ASPNG	Q39491	chlamydomon
893	5	2.6	262	1	NCAP_INSVN	P28975	impatiens n	966	5	2.6	277	1	PYRF_ASPOR	P07817	aspergillus
894	5	2.6	262	1	RNC_MYCGE	Q01808	impatiens n	967	5	2.6	278	1	CR7C_RHOSH	O13416	aspergillus
895	5	2.6	262	1	YBEM_ECO57	P47607	mycoplasma	968	5	2.6	278	1	EXSA_PSEAE	P26993	pseudomonas
896	5	2.6	262	1	YBEM_ECOLI	P58054	escherichia	969	5	2.6	278	1	OSTP_BOVIN	P31096	bos taurus
897	5	2.6	262	1	YBEM_ECOLI	P39874	escherichia	970	5	2.6	278	1	PYRF_ASPFU	O13410	aspergillus
898	5	2.6	263	1	AROE_HELPJ	P56119	helicobacte	971	5	2.6	278	1	PYRF_CLAFU	Q9Hfv8	cladosporiu
899	5	2.6	263	1	Y169_METJA	Q57633	methanococc	972	5	2.6	278	1	PYRF_PENCH	P09463	penicillium
900	5	2.6	263	1	YBGF_ECOLI	P45955	escherichia	973	5	2.6	278	1	TNF6_RAT	P36940	rattus norv
901	5	2.6	263	1	YBGF_ECOLI	P45427	escherichia	974	5	2.6	279	1	DHP1_ECOLI	P11744	escherichia
902	5	2.6	264	1	YBGF_ECOLI	P08983	xenopus lae	975	5	2.6	279	1	HRF5_HUMAN	Q96kn8	homo sapien
903	5	2.6	264	1	G3P_BORHE	P46796	borrelia he	976	5	2.6	279	1	TNF6_MOUSE	P41047	mus musculus
904	5	2.6	264	1	HRCA_CAMJE	Q9PP92	campylobact	977	5	2.6	279	1	Y43D_MYCPN	P75155	mycoplasma
905	5	2.6	264	1	RS2_HELPJ	Q9Z170	helicobacte	978	5	2.6	280	1	NADE_AERPE	P27450	arabidopsis
906	5	2.6	264	1	RS2_HELPJ	P56009	helicobacte	979	5	2.6	280	1	TNF6_CERTO	Q9Yail	aeropyrum p
907	5	2.6	264	1	RSF4_STRPU	P46771	strongyloce	980	5	2.6	280	1	TNF6_MACMU	Q9Bdn1	cercobace
908	5	2.6	264	1	WBBI_ECOLI	P36667	escherichia	981	5	2.6	280	1	Y466_HAEIN	Q9my16	macaca mula
909	5	2.6	264	1	Y557_PASMU	Q9CN81	pasteurella	982	5	2.6	280	1		P44000	haemophilus

691	1	DCX1_FOWPV	219	2.6	5	764	5	2.6	233	1	MRKB_KLEPN	P21646 klebsiella
692	1	GL22_ARATH	219	2.6	5	765	5	2.6	233	1	MSP1_PLAFD	P13827 plasmodium
693	1	GPH_RHOCA	219	2.6	5	766	5	2.6	234	1	KDSB_AQUAE	O66914 aquifex aeo
694	1	GSHE_PIG	219	2.6	5	767	5	2.6	234	1	VPYR_RHIME	Q928n8 rhizobium m
695	1	PA44_ECOLI	219	2.6	5	768	5	2.6	234	1	VSP1_AGRKH	P26324 agkistrodon
696	1	UNG_LACIA	219	2.6	5	769	5	2.6	234	1	YGS8_ECOLI	P52054 escherichia
697	1	AROD_METJA	220	2.6	5	770	5	2.6	234	1	YKFA_BACSU	O34851 bacillus su
698	1	NRG2_YEAST	220	2.6	5	771	5	2.6	235	1	L433_XENLA	O34851 xenopus lae
699	1	ROX3_YEAST	220	2.6	5	772	5	2.6	235	1	MSRA_HUMAN	Q9u168 homo sapien
700	1	YICG_HAEIN	220	2.6	5	773	5	2.6	235	1	RPSE_CLOAB	P33657 clostridium
701	1	GSXY_RAT	221	2.6	5	774	5	2.6	235	1	YIP2_YEAST	P40455 saccharomyc
702	1	GSPI_VIBCH	221	2.6	5	775	5	2.6	236	1	COAT_MCV	P11642 maize chlor
703	1	MLPE_BUCAI	221	2.6	5	776	5	2.6	236	1	RS5_PYRAB	Q9V1V5 pyrococcus
704	1	NUSB_SINY3	221	2.6	5	777	5	2.6	236	1	RS5_PYRBO	O59439 pyrococcus
705	1	SEGA_BPT4	221	2.6	5	778	5	2.6	236	1	UCRA_TOBAC	P51134 nicotiana t
706	1	SEGT_CHICK	221	2.6	5	779	5	2.6	236	1	YMI3_CAEEL	Q21018 caenorhabdi
707	1	SFR2_HUMAN	221	2.6	5	780	5	2.6	237	1	PYRF_RHILO	Q986d5 rhizobium l
708	1	SFR2_MOUSE	221	2.6	5	781	5	2.6	237	1	SPIN_HUMAN	Q9V657 homo sapien
709	1	FLGH_RHOSH	221	2.6	5	782	5	2.6	238	1	IBP6_MOUSE	P47880 mus musculu
710	1	HXC5_HUMAN	222	2.6	5	783	5	2.6	238	1	RK5_ODOSI	P49347 odontella s
711	1	HXC5_MOUSE	222	2.6	5	784	5	2.6	238	1	SFR7_HUMAN	Q16629 homo sapien
712	1	YEO7_SCHPO	222	2.6	5	785	5	2.6	239	1	OMPR_ECOLI	P03025 escherichia
713	1	CDRI_HUMAN	223	2.6	5	786	5	2.6	239	1	RP35_BACTK	P26763 bacillus th
714	1	PATM_VIBHA	223	2.6	5	787	5	2.6	239	1	RPSE_BACSU	P06222 bacillus su
715	1	RB21_CANFA	223	2.6	5	788	5	2.6	239	1	RS3_SINY3	P73314 synecocyst
716	1	UBL1_HORSE	223	2.6	5	789	5	2.6	239	1	RS3_YEAST	P05750 saccharomyc
717	1	UBL1_HUMAN	223	2.6	5	790	5	2.6	239	1	T2E8_ECOLI	P50195 escherichia
718	1	UBL1_MOUSE	223	2.6	5	791	5	2.6	239	1	Y332_MYCGE	P47574 mycoplasma
719	1	UBL1_RAT	223	2.6	5	792	5	2.6	240	1	CC14_SCHPO	P36589 schizosacch
720	1	YF94_HAEIN	223	2.6	5	793	5	2.6	240	1	LPXH_ECO57	O8XCU1 escherichia
721	1	Y209_LISMO	224	2.6	5	794	5	2.6	240	1	LPXH_ECOLI	P43341 escherichia
722	1	Y241_LISIN	224	2.6	5	795	5	2.6	240	1	SPIN_MOUSE	O61142 mus musculu
723	1	YJD7_YEAST	224	2.6	5	796	5	2.6	240	1	YD37_MYCTU	Q10647 mycobacteri
724	1	COX2_APILI	225	2.6	5	797	5	2.6	241	1	CL11_HUMAN	O00299 homo sapien
725	1	RB21_HUMAN	225	2.6	5	798	5	2.6	241	1	CL11_MOUSE	Q921q5 mus musculu
726	1	VLT3_FOWPV	225	2.6	5	799	5	2.6	241	1	YFHO_HAEIN	P44676 haemophilus
727	1	DCPR_BACSU	226	2.6	5	800	5	2.6	242	1	TRPD_BACCA	P30525 bacillus ca
728	1	IBP6_RAT	226	2.6	5	801	5	2.6	242	1	YD46_HAEIN	P44162 haemophilus
729	1	TPIS_METKA	226	2.6	5	802	5	2.6	243	1	CAVT_BRALA	P05548 branchiosto
730	1	TRPF_METTM	226	2.6	5	803	5	2.6	243	1	RS3_HUMAN	P23396 homo sapien
731	1	YQ00_LISIN	226	2.6	5	804	5	2.6	243	1	RS3_MANSE	P48153 manduca sex
732	1	YRAL_YEAST	226	2.6	5	805	5	2.6	243	1	RS3_MOUSE	P17073 mus musculu
733	1	YS21_CAEEL	226	2.6	5	806	5	2.6	244	1	BOS1_YEAST	P25385 saccharomyc
734	1	ID11_SCHPO	227	2.6	5	807	5	2.6	244	1	RS4E_SULAC	O05634 sulfolobus
735	1	PHLB_SERLI	227	2.6	5	808	5	2.6	245	1	KSU5_ECOLI	P42216 escherichia
736	1	RMW_RICPR	227	2.6	5	809	5	2.6	245	1	RS3_ICTPU	Q90Ys2 ictalurus p
737	1	CAPE_STAAU	228	2.6	5	810	5	2.6	246	1	BRM1_HUMAN	Q9HCU9 homo sapien
738	1	IFE2_CAEEL	228	2.6	5	811	5	2.6	246	1	BRM1_MOUSE	Q99420 mus musculu
739	1	LCK_AEDAE	228	2.6	5	812	5	2.6	246	1	NO25_MEDSA	P10322 medicago sa
740	1	PCAL_ACICA	228	2.6	5	813	5	2.6	246	1	RIWM_ANASP	O8YCB1 anabaena sp
741	1	TF2D_PLAFA	228	2.6	5	814	5	2.6	246	1	RS3A_XENLA	P02350 xenopus lae
742	1	VSPA_LACMU	228	2.6	5	815	5	2.6	246	1	RS3B_XENLA	P47835 xenopus lae
743	1	Y855_METJA	228	2.6	5	816	5	2.6	246	1	RS3_DROME	Q06559 drosophila
744	1	R33B_HUMAN	229	2.6	5	817	5	2.6	246	1	Y586_BUCAI	P57646 buchnera ap
745	1	R33B_MOUSE	229	2.6	5	818	5	2.6	246	1	YSW4_CAEEL	Q10019 caenorhabdi
746	1	RIBE_METTH	229	2.6	5	819	5	2.6	247	1	PCN1_AERPE	Q9YfT8 aeropyrum p
747	1	VG14_BPPH8	229	2.6	5	820	5	2.6	247	1	PCNA_HALNI	Q9P9h8 haloarcula
748	1	Y575_METJA	229	2.6	5	821	5	2.6	247	1	PCNA_HALNI	Q9Hn45 haloarcula
749	1	YT67_CAEEL	229	2.6	5	822	5	2.6	247	1	ZO20_XENLA	P18744 xenopus lae
750	1	PHOU_CAUCR	230	2.6	5	823	5	2.6	248	1	VP28_SCHPO	O13872 schizosacch
751	1	SFSA_PYRFU	230	2.6	5	824	5	2.6	248	1	Y468_BORBU	O51424 borrelia bu
752	1	Y394_PSEAE	230	2.6	5	825	5	2.6	248	1	YDFG_ECOLI	P39831 escherichia
753	1	YCAP_ECOLI	230	2.6	5	826	5	2.6	248	1	YFGE_ECOLI	P76570 escherichia
754	1	RHIA_RHILV	231	2.6	5	827	5	2.6	249	1	CISH_CHICK	Q9PW70 gallus gall
755	1	RHIA_RHILV	231	2.6	5	828	5	2.6	249	1	RS3_SCHPO	O60128 schizosacch
756	1	RUIC_YEAST	231	2.6	5	829	5	2.6	249	1	YAPL_ECOLI	Q47151 escherichia
757	1	SOML_SPAAU	231	2.6	5	830	5	2.6	250	1	GLO2_BUCAP	Q08889 buchnera ap
758	1	SOML_SPAAU	231	2.6	5	831	5	2.6	250	1	IF2B_CAEEL	Q21230 caenorhabdi
759	1	SOML_SCIOC	231	2.6	5	832	5	2.6	250	1	REP2_BUCAI	Q9ZY99 buchnera ap
760	1	SOML_SIGOU	231	2.6	5	833	5	2.6	250	1	REPA_AGRTU	P15394 agrobacteri
761	1	V054_FOWPV	231	2.6	5	834	5	2.6	250	1	YCW6_YEAST	P25659 saccharomyc
762	1	ISPD_BACSU	232	2.6	5	835	5	2.6	251	1	AMPB_BORBU	O51132 borrelia bu
763	1	SPH1_HUMAN	232	2.6	5	836	5	2.6	251	1	BIOC_ECOLI	P12999 escherichia

399	1	132	2.6	5	472	5	2.6	158	1	IF53_ARATH	Q9c505 arabidopsis
400	1	133	2.6	5	473	5	2.6	158	1	YJEE_HAEIN	P44492 haemophilus
401	1	133	2.6	5	474	5	2.6	159	1	FMI_MORBO	P20657 moraxella b
402	1	134	2.6	5	475	5	2.6	159	1	IF51_NICPL	P24921 nicotiana p
403	1	134	2.6	5	476	5	2.6	159	1	IF52_MEDSA	Q945f4 medicago sa
404	1	134	2.6	5	477	5	2.6	159	1	IF52_NICPL	P24922 nicotiana p
405	1	134	2.6	5	478	5	2.6	159	1	IF53_LYCES	Q9axq4 lycopersico
406	1	134	2.6	5	479	5	2.6	159	1	RGSD_HUMAN	Q14921 homo sapien
407	1	135	2.6	5	480	5	2.6	159	1	IF53_LYCES	P23326 spinacia ol
408	1	136	2.6	5	481	5	2.6	159	1	UK35_SPIOL	Q9c9g6 arabidopsis
409	1	136	2.6	5	482	5	2.6	159	1	UCRP_MOUSE	Q9c9g6 arabidopsis
410	1	136	2.6	5	483	5	2.6	160	1	IF52_LYCES	Q64339 mus musculus
411	1	137	2.6	5	484	5	2.6	160	1	IF53_SOLUT	Q9axq5 lycopersico
412	1	137	2.6	5	485	5	2.6	160	1	IF54_LYCES	P56335 solanum tub
413	1	138	2.6	5	486	5	2.6	160	1	IF54_LYCES	Q9axq3 lycopersico
414	1	138	2.6	5	487	5	2.6	160	1	IF54_MANES	Q9ax14 manihot esc
415	1	138	2.6	5	488	5	2.6	160	1	MD2_BOVIN	P28754 bos taurus
416	1	138	2.6	5	489	5	2.6	160	1	PHAA_AGLNE	P28755 aglaotahamni
417	1	139	2.6	5	490	5	2.6	160	1	PRU1_PUAR	O50001 prunus arme
418	1	139	2.6	5	491	5	2.6	160	1	YD49_HAEIN	P45173 haemophilus
419	1	140	2.6	5	492	5	2.6	161	1	IF51_MEDSA	P26564 medicago sa
420	1	140	2.6	5	493	5	2.6	161	1	IF52_CABEL	O20751 caenorhabdi
421	1	140	2.6	5	494	5	2.6	161	1	LE23_ARCFU	O29626 archaeoglob
422	1	140	2.6	5	495	5	2.6	161	1	YJBL_YEAST	P47076 saccharomyc
423	1	141	2.6	5	496	5	2.6	162	1	ATPE_BACCA	P41014 bacillus ca
424	1	141	2.6	5	497	5	2.6	162	1	MOTB_BPT4	Q01437 bacterioph
425	1	143	2.6	5	498	5	2.6	163	1	CYPH_BOVIN	P04374 bos taurus
426	1	143	2.6	5	499	5	2.6	163	1	DCMS_HYDPS	P19915 hydrogenoph
427	1	144	2.6	5	500	5	2.6	163	1	MLRB_CHICK	P02611 gallus gall
428	1	144	2.6	5	501	5	2.6	163	1	PHCA_ANASP	P07121 anabaena sp
429	1	145	2.6	5	502	5	2.6	164	1	UCRP_HUMAN	P05161 homo sapien
430	1	145	2.6	5	503	5	2.6	164	1	VG59_BPML5	O05272 mycobacteri
431	1	146	2.6	5	504	5	2.6	166	1	ING_BOVIN	P07353 bos taurus
432	1	146	2.6	5	505	5	2.6	166	1	ING_SHEEP	P17773 ovis aries
433	1	146	2.6	5	506	5	2.6	166	1	OV22_ONCVO	P29779 onchocerca
434	1	146	2.6	5	507	5	2.6	167	1	RB1A_DICDI	P34139 dictyosteli
435	1	146	2.6	5	508	5	2.6	167	1	YRB2_SYNP6	P23656 synchococc
436	1	146	2.6	5	509	5	2.6	168	1	AR20_SCHPO	Q23352 schizosacch
437	1	146	2.6	5	510	5	2.6	168	1	NADD_AQUAE	O66452 aquifex aeo
438	1	147	2.6	5	511	5	2.6	168	1	PUR6_ECOLI	P09028 escherichia
439	1	147	2.6	5	512	5	2.6	168	1	RBS_CHLMO	P17537 chlamydomon
440	1	147	2.6	5	513	5	2.6	169	1	GSHH_NICSY	P30708 nicotiana s
441	1	148	2.6	5	514	5	2.6	169	1	GSHH_TOBAC	O9fxs3 nicotiana t
442	1	148	2.6	5	515	5	2.6	170	1	GSHH_GOSHI	O49089 gossypium h
443	1	148	2.6	5	516	5	2.6	171	1	CYP2_CABEL	P52010 caenorhabdi
444	1	148	2.6	5	517	5	2.6	171	1	LACB_STRMU	P26424 streptococc
445	1	148	2.6	5	518	5	2.6	171	1	PHAF_AGLNE	P34812 aglaotahamni
446	1	149	2.6	5	519	5	2.6	171	1	RS13_HALNI	Q9v2w4 halobacteri
447	1	149	2.6	5	520	5	2.6	171	1	YCF3_NBPOL	Q9t102 nephroselmi
448	1	149	2.6	5	521	5	2.6	171	1	YOQ8_CABEL	P34646 caenorhabdi
449	1	149	2.6	5	522	5	2.6	171	1	YTF6_CABEL	Q19005 caenorhabdi
450	1	149	2.6	5	523	5	2.6	172	1	PKIA_DICDI	Q23921 dictyosteli
451	1	150	2.6	5	524	5	2.6	172	1	STEL_DROME	P15021 drosophila
452	1	150	2.6	5	525	5	2.6	172	1	TASM_POVBA	P15000 polyomaviru
453	1	150	2.6	5	526	5	2.6	172	1	TASM_POVBA	P03082 haemophilus
454	1	151	2.6	5	527	5	2.6	173	1	CCWE_HAEIN	P45036 haemophilus
455	1	151	2.6	5	528	5	2.6	173	1	LMIP_HUMAN	P55344 homo sapien
456	1	151	2.6	5	529	5	2.6	174	1	AXI2_ARATH	P49678 arabidopsis
457	1	152	2.6	5	530	5	2.6	174	1	OV21_ONCVO	P29778 onchocerca
458	1	152	2.6	5	531	5	2.6	174	1	VS11_ROTGI	P30890 rotavirus (
459	1	152	2.6	5	532	5	2.6	175	1	DEF_CAMJE	Q9pi18 campylobact
460	1	153	2.6	5	533	5	2.6	175	1	TFA_BPWU	Q9tlu9 bacterioph
461	1	153	2.6	5	534	5	2.6	175	1	Y433_METJA	O57875 methanococc
462	1	154	2.6	5	535	5	2.6	175	1	YXG2_YEAST	P53080 saccharomyc
463	1	154	2.6	5	536	5	2.6	176	1	UCRI_BRAJA	P51130 bradyrhizob
464	1	154	2.6	5	537	5	2.6	177	1	CCAC_CHICK	Q73707 gallus gall
465	1	155	2.6	5	538	5	2.6	177	1	RS13_HALMA	Q00861 haloarcula
466	1	155	2.6	5	539	5	2.6	178	1	COAD_ZYMMO	Q9rme4 zymomonas m
467	1	155	2.6	5	540	5	2.6	178	1	GIDB_HELPJ	O9zm61 helicobacte
468	1	156	2.6	5	541	5	2.6	178	1	GIDB_HELPJ	O25703 helicobacte
469	1	156	2.6	5	542	5	2.6	178	1	KITH_MXVLA	P28851 myxoma viru
470	1	157	2.6	5	543	5	2.6	178	1	KITH_MXVLA	Q9g86 myxoma viru
471	1	158	2.6	5	544	5	2.6	178	1	PGSA_AQUAE	O67908 aquifex aeo

253	6	3.1	2774	1	MAPA_RAT	P34926	rattus norv	326	5	2.6	107	1	RP5M_AZQVI	P17160	azotobacter
254	6	3.1	2805	1	MAPA_HUMAN	P78559	homo sapien	327	5	2.6	108	1	Y095_ARCFU	O30141	archaeoglob
255	6	3.1	3093	1	POLG_BSTV1	O65730	b genome po	328	5	2.6	108	1	YB01_SYNV3	P72745	synechocyst
256	6	3.1	3256	1	KI67_HUMAN	P46013	homo sapien	329	5	2.6	108	1	YD88_SCHPO	Q10413	schizosacch
257	6	3.1	3911	1	AKA9_HUMAN	Q99966	h a-kinase	330	5	2.6	109	1	CL11_PIG	Q29238	su scrofa
258	5	2.6	15	1	UBL1_MONDO	P50103	monodelphis	331	5	2.6	109	1	KV3B_HUMAN	P01620	homo sapien
259	5	2.6	20	1	LPP2_HUMAN	P56642	homo sapien	332	5	2.6	109	1	KV3D_HUMAN	P01622	homo sapien
260	5	2.6	26	1	UBL1_BOVIN	P23356	bos taurus	333	5	2.6	109	1	KV3E_HUMAN	P01623	homo sapien
261	5	2.6	34	1	EM1_ENSMI	P27205	ensis minor	334	5	2.6	109	1	KV3G_HUMAN	P04206	homo sapien
262	5	2.6	50	1	YF24_HAEIN	P44244	haemophilus	335	5	2.6	109	1	Y265_UREPA	Q9pqm6	ureaplasma
263	5	2.6	54	1	RK32_TOBAC	P12198	nicotiana t	336	5	2.6	109	1	YJ49_YEAST	P47126	saccharomyc
264	5	2.6	55	1	A70A_DROME	O18666	drosophila	337	5	2.6	110	1	YG12_BPPF1	P25132	bacterioph
265	5	2.6	55	1	A70A_DROME	P05623	drosophila	338	5	2.6	110	1	YCKD_BACSU	P42402	bacillus su
266	5	2.6	55	1	A70A_DROSE	O18417	drosophila	339	5	2.6	112	1	FMDB_METME	Q50229	methylophil
267	5	2.6	60	1	HMA2_HELTR	P17138	helobdella	340	5	2.6	114	1	MYH8_MOUSE	P13542	mus musculus
268	5	2.6	62	1	RL28_LISMO	Q92a12	listeria mo	341	5	2.6	114	1	RL19_BACHD	Q9ka16	bacillus ha
269	5	2.6	62	1	GBAK_CRIGR	Q60397	cricetulus	342	5	2.6	116	1	RL31_CHLRE	P45841	chlamydomon
270	5	2.6	64	1	RK35_CYACA	Q9tir9	cyamidium c	343	5	2.6	116	1	RR8_MUSAC	Q42362	musa acumin
271	5	2.6	70	1	Y4ON_RHISN	P55599	rhizobium s	344	5	2.6	116	1	YC20_MESVI	Q9mul5	mesostigma
272	5	2.6	72	1	RL15_BACLI	P35138	bacillus li	345	5	2.6	116	1	YG10_YEAST	P53222	saccharomyc
273	5	2.6	73	1	RL15_STACA	P35139	staphylococ	346	5	2.6	116	1	YJG7_YEAST	P40365	saccharomyc
274	5	2.6	76	1	SECG_BACSU	Q32233	bacillus su	347	5	2.6	118	1	YAB7_SCHPO	Q09774	schizosacch
275	5	2.6	76	1	UBIQ_ACACA	P49634	acanthamoeb	348	5	2.6	119	1	ACPS_STAAH	Q99S14	staphylococ
276	5	2.6	76	1	UBIQ_ACECL	P42739	acetabulari	349	5	2.6	119	1	ACPS_STAAU	Q9zah6	staphylococ
277	5	2.6	76	1	UBIQ_AGLNE	P42740	aglaethami	350	5	2.6	120	1	YF14_MYCPN	P75272	mycoplasma
278	5	2.6	76	1	UBIQ_CABEL	P14792	caenorhabdi	351	5	2.6	120	1	YQ58_MYCTU	P71955	mycobacteri
279	5	2.6	76	1	UBIQ_CHLRE	P14624	chlamydomon	352	5	2.6	121	1	RL14_AQUAE	O67570	aquifex aeo
280	5	2.6	76	1	UBIQ_COPCO	P19848	coprinus co	353	5	2.6	121	1	RL14_AQUFY	Q92i42	aquifex pyr
281	5	2.6	76	1	UBIQ_DICDI	P08618	dicrosteli	354	5	2.6	121	1	VNBP_POPMV	Q02123	poplar mosa
282	5	2.6	76	1	UBIQ_DROME	Q9vz14	drosophila	355	5	2.6	121	1	YIF9_YEAST	Q40520	saccharomyc
283	5	2.6	76	1	UBIQ_EIMBO	P46574	eimeria bov	356	5	2.6	122	1	YIF9_YEAST	P06381	marchantia
284	5	2.6	76	1	UBIQ_EUPEU	P23324	euplotes eu	357	5	2.6	122	1	RL14_MYCTU	P95068	mycobacteri
285	5	2.6	76	1	UBIQ_HUMAN	P02248	homo sapien	358	5	2.6	122	1	YV66_METTF	P29580	methanobact
286	5	2.6	76	1	UBIQ_LEIMA	Q05550	leishmania	359	5	2.6	122	1	YV66_METTF	P29581	methanobact
287	5	2.6	76	1	UBIQ_LEITA	P49635	leishmania	360	5	2.6	123	1	RS12_CHLPN	Q92800	chlamydia p
288	5	2.6	76	1	UBIQ_NEUCR	P13117	neurospora	361	5	2.6	123	1	UL59_RCMVA	P16746	human cytom
289	5	2.6	76	1	UBIQ_PHYIN	P22589	phytophthor	362	5	2.6	124	1	SOR_PYRFU	P82385	pyrococcus
290	5	2.6	76	1	UBIQ_SOYBN	P03993	glycine max	363	5	2.6	124	1	SPH2_SULSO	Q9uwy8	sulfolobus
291	5	2.6	76	1	UBIQ_STRPU	P23398	strongyloce	364	5	2.6	124	1	YCX3_CHLRE	P05724	chlamydomon
292	5	2.6	76	1	UBIQ_TETPY	P20685	tetrahymena	365	5	2.6	125	1	ALR_MOUSE	P56213	mus musculus
293	5	2.6	76	1	UBIQ_TRYBB	P15174	trypanosoma	366	5	2.6	125	1	ALR_MOUSE	Q93042	rattus norv
294	5	2.6	76	1	UBIQ_TRYCR	P08565	trypanosoma	367	5	2.6	125	1	H2BH_HUMAN	Q93078	homo sapien
295	5	2.6	76	1	UBIQ_YEAST	P04838	saccharomyc	368	5	2.6	125	1	PYP_CHRSA	P81046	chromatium
296	5	2.6	77	1	UBIL_NPVAC	P16709	autographa	369	5	2.6	125	1	RPAG_YEAST	P32529	saccharomyc
297	5	2.6	82	1	Y235_METJA	Q57687	methanococ	370	5	2.6	125	1	Y364_AQUAE	O56689	aquifex aeo
298	5	2.6	83	1	PYS1_SYNV3	P73202	synechocyst	371	5	2.6	125	1	YCX1_GUTHI	O78420	guillardia
299	5	2.6	84	1	ACPH_MYCGE	P47529	mycoplasma	372	5	2.6	125	1	YEGR_ECOLI	P76406	escherichia
300	5	2.6	84	1	YCX9_OENHO	Q9mtm6	oenothera h	373	5	2.6	126	1	PAND_YERPE	Q8zbn6	versinia pe
301	5	2.6	87	1	VAPL_BP186	P21681	bacterioph	374	5	2.6	126	1	YK50_ARCFU	O28229	archaeoglob
302	5	2.6	87	1	VE4_HPV51	P26548	human papil	375	5	2.6	127	1	RS8E_PYRHO	O58830	pyrococcus
303	5	2.6	90	1	RPOL_SULAC	P46217	sulfolobus	376	5	2.6	127	1	RT08_ACACA	P46757	acanthamoeb
304	5	2.6	91	1	SR19_METTH	O26267	methanobact	377	5	2.6	127	1	RT12_CHOCR	P48858	chondrus cr
305	5	2.6	92	1	RR19_MALZE	P06588	zea mays (m	378	5	2.6	127	1	TRM1_ECOLI	P10026	escherichia
306	5	2.6	92	1	RR19_ORYSA	P12153	oryza sativ	379	5	2.6	127	1	TRM6_ECOLI	P18808	escherichia
307	5	2.6	92	1	RR19_SECEC	P33954	secale cere	380	5	2.6	127	1	TRM7_ECOLI	P33787	escherichia
308	5	2.6	93	1	SP3D_BACSU	P15281	bacillus su	381	5	2.6	127	1	TRM9_ECOLI	P18807	escherichia
309	5	2.6	93	1	UBIL_NPVOP	Q05120	orgyia pseu	382	5	2.6	128	1	KV3K_HUMAN	P06311	homo sapien
310	5	2.6	96	1	VG26_BPMU	Q9t148	bacterioph	383	5	2.6	128	1	Y097_BORBU	O51124	borrelia bu
311	5	2.6	96	1	VPR_SIVCZ	P17287	chimpanzee	384	5	2.6	128	1	Y358_METJA	Q37804	methanococ
312	5	2.6	96	1	Y448_AQUAE	O66756	aquifex aeo	385	5	2.6	128	1	Y582_METJA	Q38002	methanococ
313	5	2.6	96	1	YEBG_ECOLI	P33220	escherichia	386	5	2.6	128	1	Y040_BACAN	Q9rmz2	bacillus su
314	5	2.6	97	1	CU83_HSV62	P52461	human herpe	387	5	2.6	129	1	PLGB_BACSU	P24500	bacillus su
315	5	2.6	97	1	RL28_BRUME	Q8yjm6	brucella me	388	5	2.6	129	1	GSPL_PSPAE	Q00516	pseudomonas
316	5	2.6	98	1	ACYO_CHICK	P07032	gallus gall	389	5	2.6	129	1	H3_LEIIN	P40285	leishmania
317	5	2.6	98	1	NULM_HIFAM	Q9zyy3	hippopotamu	390	5	2.6	129	1	HYTA_APIWE	Q10416	apis mellif
318	5	2.6	98	1	VG04_BPMD2	O64200	mycobacteri	391	5	2.6	129	1	KV3L_HUMAN	P18135	homo sapien
319	5	2.6	99	1	ILB6_CABEL	P56173	caenorhabdi	392	5	2.6	129	1	KV3M_HUMAN	P18136	homo sapien
320	5	2.6	99	1	PSD2_BOVIN	P56701	bos taurus	393	5	2.6	129	1	YG2Z_YEAST	P53268	saccharomyc
321	5	2.6	100	1	KV3C_HUMAN	P01621	homo sapien	394	5	2.6	130	1	YQ26_BACAN	Q9rn06	bacillus an
322	5	2.6	102	1	YEG5_YEAST	P43539	saccharomyc	395	5	2.6	131	1	PSBU_SYNV3	Q55332	synechocyst
323	5	2.6	105	1	NDCL_BRAJA	P26024	bradyrhizob	396	5	2.6	131	1	RS12_DEIRA	Q3rxk7	deinococcus
324	5	2.6	107	1	R520_THETH	P80380	thermus the	397	5	2.6	131	1	YC85_METTH	P27346	methanobact
325	5	2.6	107	1	E312_ADE05	P06496	human adeno	398	5	2.6	132	1	RS8_BORBU	O51445	borrelia bu

107	1	402	3.1	6	180	6	3.1	690	1	PCD8_DROME	Q9vq79 drosophila
108	1	403	3.1	6	181	6	3.1	692	1	YFB3_YEAST	P43579 saccharomyc
109	1	406	3.1	6	182	6	3.1	710	1	Y026_YEAST	P39744 saccharomyc
110	1	407	3.1	6	183	6	3.1	717	1	MCCA_MOUSE	Q99mr8 mus musculus
111	1	425	3.1	6	184	6	3.1	727	1	CTCF_HUMAN	P49711 homo sapien
112	1	425	3.1	6	185	6	3.1	728	1	CTCF_CHICK	Q08705 gallus gall
113	1	425	3.1	6	186	6	3.1	733	1	K6A2_MOUSE	Q9wtu3 mus musculus
114	1	425	3.1	6	187	6	3.1	752	1	PHSM_STRPN	P29849 streptococc
115	1	426	3.1	6	188	6	3.1	768	1	CNIC_RAT	Q63421 rattus norv
116	1	431	3.1	6	189	6	3.1	790	1	CB80_HUMAN	Q09161 homo sapien
117	1	439	3.1	6	190	6	3.1	805	1	CC48_CAPAN	Q96372 capsicum an
118	1	441	3.1	6	191	6	3.1	806	1	MK07_MOUSE	Q9wvs8 mus musculus
119	1	441	3.1	6	192	6	3.1	807	1	CC48_SOYBN	Q94774 glycine max
120	1	443	3.1	6	193	6	3.1	809	1	CC48_ARATH	P54609 arabidopsis
121	1	444	3.1	6	194	6	3.1	809	1	HRPB_ECOLI	P37024 escherichia
122	1	449	3.1	6	195	6	3.1	815	1	KY07_HUMAN	Q13164 homo sapien
123	1	449	3.1	6	196	6	3.1	827	1	GYRA_HELPY	P48370 helicobacte
124	1	457	3.1	6	197	6	3.1	834	1	PLSB_PSEAE	Q9hbw7 pseudomonas
125	1	461	3.1	6	198	6	3.1	835	1	CC48_YEAST	P25694 saccharomyc
126	1	463	3.1	6	199	6	3.1	844	1	PHSG_DROME	Q9xt19 drosophila
127	1	465	3.1	6	200	6	3.1	855	1	NOL1_HUMAN	P46087 homo sapien
128	1	467	3.1	6	201	6	3.1	858	1	CNRA_BOVIN	P11541 bos taurus
129	1	471	3.1	6	202	6	3.1	858	1	CNRA_MOUSE	P27664 mus musculus
130	1	475	3.1	6	203	6	3.1	859	1	CNRA_HUMAN	P16499 homo sapien
131	1	475	3.1	6	204	6	3.1	860	1	CNRA_CANFA	Q28463 canis famli
132	1	476	3.1	6	205	6	3.1	863	1	APCE_GALSU	P35911 galleria s
133	1	478	3.1	6	206	6	3.1	864	1	YG3M_YEAST	P48337 saccharomyc
134	1	481	3.1	6	207	6	3.1	875	1	HELX_SULSO	P95949 sulfolobus
135	1	484	3.1	6	208	6	3.1	882	1	DPO1_SULSO	P26811 sulfolobus
136	1	484	3.1	6	209	6	3.1	899	1	YMJ3_YEAST	Q04500 saccharomyc
137	1	485	3.1	6	210	6	3.1	906	1	FOX2_CANTR	P22414 candida tro
138	1	489	3.1	6	211	6	3.1	924	1	DPOL_BPSPI	P23114 bacterioph
139	1	494	3.1	6	212	6	3.1	928	1	RR44_HUMAN	Q9Y211 homo sapien
140	1	496	3.1	6	213	6	3.1	960	1	DLG1_DROME	P31007 drosophila
141	1	496	3.1	6	214	6	3.1	994	1	DPO2_KLULA	P05468 kluyveromyc
142	1	501	3.1	6	215	6	3.1	994	1	SVV_XYLFA	Q9ph12 xyliella fas
143	1	501	3.1	6	216	6	3.1	1011	1	VPS3_YEAST	P23643 saccharomyc
144	1	503	3.1	6	217	6	3.1	1021	1	MAFA_MOUSE	Q9qyr6 mus musculus
145	1	506	3.1	6	218	6	3.1	1053	1	SPS_SOLTU	Q43845 solanum tub
146	1	508	3.1	6	219	6	3.1	1057	1	ANPA_MOUSE	P18293 mus musculus
147	1	514	3.1	6	220	6	3.1	1057	1	ANPA_RAT	P18910 rattus norv
148	1	514	3.1	6	221	6	3.1	1061	1	ANPA_HUMAN	P16066 homo sapien
149	1	515	3.1	6	222	6	3.1	1084	1	XP01_YEAST	P30822 saccharomyc
150	1	516	3.1	6	223	6	3.1	1091	1	JSN1_YEAST	P47135 saccharomyc
151	1	519	3.1	6	224	6	3.1	1102	1	MYSC_CHICK	P29616 gallus gall
152	1	525	3.1	6	225	6	3.1	1113	1	N116_YEAST	Q02630 saccharomyc
153	1	529	3.1	6	226	6	3.1	1119	1	RPOB_THEAQ	Q9kwu7 thermus aqu
154	1	530	3.1	6	227	6	3.1	1176	1	HMDH_PHYBL	Q12649 phycomyces
155	1	548	3.1	6	228	6	3.1	1210	1	AF4_HUMAN	P51825 homo sapien
156	1	549	3.1	6	229	6	3.1	1220	1	ATB3_HUMAN	Q16720 homo sapien
157	1	549	3.1	6	230	6	3.1	1258	1	ATB3_RAT	Q64568 rattus norv
158	1	549	3.1	6	231	6	3.1	1335	1	DP3A_DEIRA	Q9rx08 deinococcus
159	1	553	3.1	6	232	6	3.1	1344	1	IF3A_MOUSE	P23116 mus musculus
160	1	555	3.1	6	233	6	3.1	1403	1	YGN1_YEAST	Q91293 rana catesb
161	1	556	3.1	6	234	6	3.1	1407	1	TRHY_RABIT	P37709 saccharomyc
162	1	565	3.1	6	235	6	3.1	1476	1	AT7A_CRIGR	P49015 cricetus
163	1	568	3.1	6	236	6	3.1	1477	1	YORI_YEAST	P53049 saccharomyc
164	1	569	3.1	6	237	6	3.1	1489	1	YGP0_YEAST	P53115 saccharomyc
165	1	576	3.1	6	238	6	3.1	1491	1	AT7A_MOUSE	Q64430 mus musculus
166	1	577	3.1	6	239	6	3.1	1496	1	CPSM_RANCA	Q91293 ovis aries
167	1	599	3.1	6	240	6	3.1	1549	1	TRHY_SHEEP	P22793 ovis aries
168	1	600	3.1	6	241	6	3.1	1564	1	PDRA_YEAST	P51533 saccharomyc
169	1	603	3.1	6	242	6	3.1	1653	1	CLH_YEAST	P22137 saccharomyc
170	1	620	3.1	6	243	6	3.1	1850	1	VIT2_CHICK	P02845 gallus gall
171	1	620	3.1	6	244	6	3.1	1898	1	TRHY_HUMAN	Q07283 homo sapien
172	1	620	3.1	6	245	6	3.1	1906	1	YFAQ_ANASP	Q8ym40 anabaena sp
173	1	621	3.1	6	246	6	3.1	1939	1	MYHG_MESAU	P13539 mesocricetu
174	1	622	3.1	6	247	6	3.1	2052	1	FYV1_MOUSE	Q921t6 mus musculus
175	1	639	3.1	6	248	6	3.1	2349	1	TPR_HUMAN	P12270 homo sapien
176	1	642	3.1	6	249	6	3.1	2363	1	SPCO_MOUSE	Q62261 mus musculus
177	1	645	3.1	6	250	6	3.1	2364	1	SPCO_HUMAN	Q01082 homo sapien
178	1	659	3.1	6	251	6	3.1	2388	1	SPCP_RAT	Q9qwn8 rattus norv
179	1	689	3.1	6	252	6	3.1	2390	1	SPCP_HUMAN	Q15020 homo sapien

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2003, 09:55:13 ; Search time 13 seconds
(without alignments)

612.573 Million cell updates/sec

Title: US-09-868-352-23

Perfect score: 192

Sequence: 1 MRKEVPEMLNKNYKPGPQF.....EEODKEMTSKQHLFFVRKN 192

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	4.2	733	1 K6A2_HUMAN	Q15349 homo sapien
2	7	3.6	57	1 RL32_MYCTU	P58287 mycobacteri
3	7	3.6	81	1 SFR6_RABIT	O18776 oryctolagus
4	7	3.6	124	1 RL12_PORPU	P51289 porphyra pu
5	7	3.6	186	1 TATB_HAEIN	P57047 haemophilus
6	7	3.6	200	1 RACH_DICDI	Q59pr7 dictyosteli
7	7	3.6	344	1 SFR6_HUMAN	Q13247 homo sapien
8	7	3.6	381	1 DHPH_BACSH	P23307 bacillus sp
9	7	3.6	382	1 NCAP_CVPPU	P05991 porcine tra
10	7	3.6	382	1 NCAP_CVPPU	P04134 porcine tra
11	7	3.6	382	1 NCAP_CVPR8	P33463 porcine res
12	7	3.6	382	1 NCAP_CVPRM	P24411 porcine res
13	7	3.6	577	1 YG88_YEAST	P53075 saccharomyc
14	7	3.6	632	1 PTMA_BUCAI	P57635 buchnera ap
15	7	3.6	788	1 KDGG_MOUSE	Q91wg7 mus musculu
16	7	3.6	788	1 KDGG_RAT	P49620 rattus norv
17	7	3.6	791	1 KDGG_HUMAN	P49619 homo sapien
18	7	3.6	798	1 L100_ADEGX	P36856 avian adeno
19	7	3.6	859	1 SYL_BUCAI	P57519 buchnera ap
20	7	3.6	1233	1 IRS1_MOUSE	P35569 mus musculu
21	7	3.6	1242	1 IRS1_HUMAN	P35568 homo sapien
22	7	3.6	1318	1 VIVD_BPT7	P03726 bacterioph
23	7	3.6	1321	1 IRS2_MOUSE	P81122 mus musculu
24	7	3.6	1324	1 IRS2_HUMAN	Q9y4h2 homo sapien
25	7	3.6	2145	1 U520_CAEEL	Q0u290 caenorhabdi
26	7	3.6	2386	1 RAD3_SCHPO	Q02099 schizosacch
27	6	3.1	35	1 PH1L_MYCTU	P35422 mytilus cal
28	6	3.1	64	1 YM83_MYCTU	Q50682 mycobacteri
29	6	3.1	75	1 R14E_METH	O26139 methanobact
30	6	3.1	91	1 PH1L_MYTED	Q04621 mytilus edu
31	6	3.1	103	1 CYAY_RICCN	Q921h7 rickettsia
32	6	3.1	118	1 Y781_SULSO	Q9uxg8 sulfolobus
33	6	3.1	120	1 YIBI_ECOLI	P23108 escherichia

34	6	3.1	122	1 YLB9_CAEEL	P46583 caenorhabdi
35	6	3.1	125	1 H2B1_XENLA	P02281 xenopus lae
36	6	3.1	125	1 H2B2_XENLA	P06900 xenopus lae
37	6	3.1	126	1 GP48_BPSF1	O48402 bacterioph
38	6	3.1	126	1 RK12_CYAPA	P48124 cyanophora
39	6	3.1	148	1 NDK_THEAC	Q9hj59 thermoplas
40	6	3.1	149	1 RL13_CHLPM	Q928t7 chlamydia p
41	6	3.1	150	1 RL13_CHLPM	Q9pkf3 chlamydia m
42	6	3.1	150	1 RL13_CHLTR	O84127 chlamydia t
43	6	3.1	150	1 YDPC_SCHPO	Q14015 schizosacch
44	6	3.1	155	1 REG1_PYRAB	Q9v2m0 pyrococcus
45	6	3.1	155	1 REG1_PYRHO	P55685 pyrococcus
46	6	3.1	181	1 Y4WG_RHISN	P55685 rhizobium s
47	6	3.1	187	1 RK24_TOBAC	Q02764 nicotiana t
48	6	3.1	188	1 NRFX_PASMU	Q9cpk9 pasteurella
49	6	3.1	191	1 Y948_METJA	Q58358 methanococc
50	6	3.1	194	1 RS7_FUGRU	P50894 fugu rubrip
51	6	3.1	202	1 RR4_MARPO	P06358 marchantia
52	6	3.1	205	1 RR4_EUGGR	P27418 euglena gra
53	6	3.1	206	1 Y140_METJA	Q57605 methanococc
54	6	3.1	210	1 GTS1_CAEEL	Q09607 caenorhabdi
55	6	3.1	210	1 THIE_ARCFU	Q28205 archaeoglob
56	6	3.1	215	1 ERD2_PETHY	Q9zttn2 petunia hyb
57	6	3.1	221	1 YR75_CAEEL	Q09394 caenorhabdi
58	6	3.1	232	1 SLT2_CAEEL	Q19892 caenorhabdi
59	6	3.1	249	1 KTXG_KIULA	P09807 kluyveromyc
60	6	3.1	259	1 Y068_CHLTR	O84071 chlamydia t
61	6	3.1	261	1 YEV6_YEAST	P40078 saccharomyc
62	6	3.1	271	1 CEAM_ECOLI	P05820 escherichia
63	6	3.1	280	1 LBX1_HUMAN	P52954 homo sapien
64	6	3.1	280	1 LBX1_MOUSE	P52955 mus musculu
65	6	3.1	280	1 LGTE_NEIMB	Q51117 neisseria m
66	6	3.1	280	1 YB27_METJA	O58527 methanococc
67	6	3.1	289	1 GNPI_HUMAN	P46926 homo sapien
68	6	3.1	292	1 YQ14_CAEEL	Q09505 caenorhabdi
69	6	3.1	294	1 DEH1_MORSP	Q01398 moraxella s
70	6	3.1	294	1 YQ19_CAEEL	Q09507 caenorhabdi
71	6	3.1	301	1 ORT2_HUMAN	Q9bx12 homo sapien
72	6	3.1	308	1 JUNB_CYPCA	P79703 cyprinus ca
73	6	3.1	312	1 O1F1_HUMAN	O43749 homo sapien
74	6	3.1	314	1 ARCC_LACSK	O53090 lactobacill
75	6	3.1	323	1 JUND_CHICK	P27921 gallus gall
76	6	3.1	326	1 RNHL_SCHPO	Q10236 schizosacch
77	6	3.1	327	1 E13D_HORVU	Q02437 hordeum vul
78	6	3.1	330	1 YN77_YEAST	P50947 saccharomyc
79	6	3.1	331	1 YZ44_METJA	O50299 methanococc
80	6	3.1	332	1 BIOB_BACSH	P19206 bacillus sp
81	6	3.1	341	1 JUND_MOUSE	P15066 mus musculu
82	6	3.1	341	1 JUND_RAT	P52909 rattus norv
83	6	3.1	344	1 Y182_ARCFU	Q30055 archaeoglob
84	6	3.1	345	1 PRIM_PYRAB	Q9v292 pyrococcus
85	6	3.1	347	1 JUND_HUMAN	P17535 homo sapien
86	6	3.1	353	1 GB11_CAVPO	P38401 cavia porce
87	6	3.1	353	1 GB11_CHICK	P50146 gallus gall
88	6	3.1	353	1 GB11_HUMAN	P04898 homo sapien
89	6	3.1	353	1 GB11_ORYLA	P87383 oryzaes lat
90	6	3.1	353	1 GB11_RAT	P10824 rattus norv
91	6	3.1	353	1 GB11_XENLA	P27044 xenopus lae
92	6	3.1	353	1 PLSA_AGR5	Q8ug63 agrobacteri
93	6	3.1	354	1 GB12_CANFA	P38400 canis famli
94	6	3.1	354	1 GB12_CAVPO	P38402 cavia porce
95	6	3.1	354	1 GB12_CHICK	P50147 gallus gall
96	6	3.1	354	1 GB12_HUMAN	P04899 homo sapien
97	6	3.1	354	1 GB12_MOUSE	P08752 mus musculu
98	6	3.1	354	1 GB12_RAT	P04897 rattus norv
99	6	3.1	356	1 QUEA_RALSO	Q8xv55 talstonia s
100	6	3.1	360	1 CWLL_BACLI	P36550 bacillus li
101	6	3.1	360	1 FLHB_BACSU	P35538 bacillus su
102	6	3.1	370	1 CYSA_MARPO	P10091 marchantia
103	6	3.1	373	1 DP3B_MYCPU	P584r6 mycoplasma
104	6	3.1	381	1 YNU6_CAEEL	P50444 caenorhabdi
105	6	3.1	391	1 IF43_NICPL	P41380 nicotiana p
106	6	3.1	393	1 VN53_ROTPO	Q00033 porcine rot

THIS PAGE BLANK (USPTO)

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 RSKSOSS 138
|||||||
Db 267 RSKSOSS 273

RESULT 40

AAR67708

ID AAR67708 standard; Protein; 1243 AA.

XX

AC AAR67708;

XX

DT 20-JUL-1995 (first entry)

XX

DE Insulin receptor substrate-1.

XX

KW Insulin receptor substrate-1; IRS-1;

XX

KW non-insulin-dependent diabetes; NIDDM.

XX

OS Homo sapiens.

XX

PN WO9429345-A.

XX

PD 22-DEC-1994.

XX

PF 10-JUN-1994; 94WO-DK00227.

XX

PR 10-JUN-1993; 93DK-0000683.

XX

PR 09-AUG-1993; 93DK-0000915.

XX

PA (NOVO) NOVO-NORDISK AS.

XX

PI Bjoerbaek C, Frederiksen KA, Pedersen O;

XX

DR WPI; 1995-036405/05.

XX

DR N-PSDB; AAQ75977.

XX

PT DNA encoding mutated insulin receptor substrate 1 - and related
PT vectors, transformed cells and mutant proteins, useful for
PT detecting predisposition to non-insulin-dependent diabetes

XX

PS Disclosure; Page 31-38; 58pp; English.

XX

CC Genomic DNA from human leukocyte nuclei was subjected to PCR
CC amplification using primers (given in AAQ75978-Q76015) specific for the
CC human IRS-1 gene (AAQ75977). Amplified DNA was subjected to SSCP
CC analysis to detect mutations in the IRS-1 gene, indicating a
CC predisposition to NIDDM.

XX

SQ Sequence 1243 AA;

Query Match 3.6%; Score 7; DB 16; Length 1243;

Best Local Similarity 100.0%; Pred. No. 2e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 RSKSOSS 138

|||||||

Db 268 RSKSOSS 274

Search completed: February 26, 2003, 10:00:07

Job time : 66 secs

PS Claim 4; Fig 16; 86pp; English.

XX The invention relates to modulators that affect the interaction of a

CC polypeptide defining a PI (phosphotyrosine interaction) domain, or its

CC conserved variants or fragments, with APP (amyloid precursor protein).

CC The polypeptide defining a PI domain binds to the cytoplasmic domain of

CC APP, regulating the secretion of APP fragments. By modulating this

CC interaction, the modulator can inhibit or prevent APP processing and

CC trafficking, and thus progress or onset of Alzheimer's disease. The

CC modulators are used to treat and/or prevent neurodegenerative diseases

CC in mammals, particularly Alzheimer's disease and dementia. Sequences

CC AAV13448-461 represent examples of polypeptides defining a PI domain.

XX

SQ Sequence 1242 AA;

Query Match 3.6%; Score 7; DB 20; Length 1242;

Best Local Similarity 100.0%; Pred. No. 2e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 RSKSQSS 138

Db 267 RSKSQSS 273

IIIIII

RESULT 38

AAW93972

ID AAW93972 standard; protein; 1242 AA.

XX

AC AAW93972;

XX

XX

DT 30-JUN-1999 (first entry)

XX

DE Human IRS-1 and IRS-2 binding inhibitor protein.

XX

XX Insulin receptor substrate; IRS-1; IRS-2; inhibitor; disease; obesity;

KW insulin resistance; 14-3-3 protein; inhibitor; screening; diabetes;

KW diabetic retinopathy; diabetic neuropathy; glucose tolerance;

KW diabetic nephropathy; hyperinsulinaemia; hyperlipemia; arteriosclerosis;

KW hypertension; ischaemic heart disease; ischaemic brain disease;

KW peripheral embolism.

XX

OS Homo sapiens.

XX

XX WO9916462-A1.

PN

XX

PD 08-APR-1999.

XX

XX 25-SEP-1998; 98WO-JP04293.

PF

XX

XX 29-SEP-1997; 97JP-0263719.

PR

XX

XX (DAUC) DAIICHI PHARM CO LTD.

PA

XX

XX Asano T, Kanda A, Kubo H, Yazaki Y;

PI

XX

XX WPI; 1999-254929/21.

DR

XX

XX Treatment of insulin resistance using insulin receptor substrate

PT binding inhibitor

XX

XX Disclosure; Page 16-22; 30pp; Japanese.

PS

XX This invention describes a method for the treatment of diseases involving

CC insulin resistance using a substance which inhibits the binding of

CC insulin receptor substrate to 14-3-3 protein, identified by screening

CC potential inhibitors for their ability to inhibit this binding. The

CC composition described in the invention for the treatment of diseases

CC involving insulin resistance contains as active component, an inhibitor

CC of the binding of all or part of insulin receptor substrate 1 or 2

CC (IRS-1, IRS-2) to all or part of 14-3-3 protein. Disorders involving

CC insulin resistance include diabetes, diabetic retinopathy, diabetic

CC neuropathy, impaired glucose tolerance, diabetic nephropathy, hyperinsulinaemia, hyperlipemia, arteriosclerosis, hypertension,

CC obesity, ischaemic heart disease, ischaemic brain disease and peripheral

XX embolism.

XX

SQ Sequence 1242 AA;

Query Match 3.6%; Score 7; DB 20; Length 1242;

Best Local Similarity 100.0%; Pred. No. 2e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 RSKSQSS 138

Db 267 RSKSQSS 273

IIIIII

RESULT 39

AAB83921

ID AAB83921 standard; Protein; 1242 AA.

XX

AC AAB83921;

XX

XX 23-JUL-2001 (first entry)

DT

XX

DE Amino acid sequence of a human insulin receptor substrate (IRS).

XX

XX Epidermal growth factor-like domain; EGF-like domain; cancer;

KW human aspartyl beta-hydroxylase; HAAH; malignant neoplasm; tumour;

KW insulin receptor substrate; IRS.

XX

OS Homo sapiens.

XX

XX WO200135102-A2.

PN

XX

PD 17-MAY-2001.

XX

XX 08-NOV-2000; 2000WO-US30738.

PF

XX

XX 08-NOV-1999; 99US-0436184.

PR

XX

XX (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.

PA

XX

XX Wands JR, De La Monte SM, Ince N, Carlson RI;

PI

XX

XX WPI; 2001-329171/34.

DR

XX

XX N-PSDB; AAF89816.

DR

XX

XX Diagnosing malignant neoplasm in a mammal, involves contacting

PT mammalian sample with antibody that binds to human aspartyl

PT beta-hydroxylase polypeptide to form antigen-antibody complex and

PT detecting the complex

XX

XX Example 7; Page 52; 76pp; English.

PS

XX The present sequence represents a human insulin receptor substrate (IRS).

CC Compounds which inhibit IRS also inhibit growth of tumours which

CC overexpress human aspartyl (asparaginyl) beta-hydroxylase (HAAH) enzyme.

CC Epidermal growth factor (EGF)-like domains of polypeptides are

CC hydroxylated by HAAH enzymes. HAAH is used in the method of the

CC invention. The specification describes a method for diagnosing a

CC malignant neoplasm in a mammal. The method comprises contacting a body

CC fluid with an antibody which binds to HAAH polypeptide under complex

CC forming conditions, and detecting the antigen-antibody complex. The

CC method is useful for diagnosing and prognosing a malignant neoplasm in

CC a bodily fluid e.g. central nervous system (CNS)-derived body fluid,

CC blood, serum, urine, saliva, sputum, lung effusion, and ascites fluid

CC of mammal, where the neoplasm is derived from endodermal tissue and is

CC selected from colon cancer, breast cancer, pancreatic cancer, liver

CC cancer, cancer of bile ducts, and cancer of the CNS. HAAH antibodies,

CC linked to a cytotoxic agent, are useful for killing tumour cells.

XX

XX Sequence 1242 AA;

SQ

Query Match 3.6%; Score 7; DB 22; Length 1242;

Best Local Similarity 100.0%; Pred. No. 2e+02;

PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 3.6%; Score 7; DB 21; Length 823;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 182 AKOHLUF 188
| | | | |
Db 27 AKOHLUF 33

RESULT 36
AAB42816
ID AAB42816 standard; Protein; 849 AA.

XX AAB42816;
XX
DT 08-FEB-2001 (first entry)
XX

DE Human ORFX ORF2580 polypeptide sequence SEQ ID NO:5160.

XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnery; antiparotiatric; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.

XX Homo sapiens.
XX
XX WO200058473-A2.
XX
XX 05-OCT-2000.
XX
XX 31-MAR-2000; 2000WO-US086621.
XX
XX 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.

XX (CURA-) CURAGEN CORP.
XX
XX Shimkets RA, Leach M;
XX
XX WPI; 2000-602362/57.
DR N-PSDB; AAC77025.

XX

PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
PS Claim 11; Page 4342-4344; 5507pp; English.

XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC antiparotiatric; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antithyroid; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 849 AA;

Query Match 3.6%; Score 7; DB 21; Length 849;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 DWGNEQL 72
| | | | |
Db 33 DWGNEQL 39

RESULT 37
AAV13461
ID AAV13461 standard; protein; 1242 AA.

XX AAV13461;
XX
XX 26-JUL-1999 (first entry)
XX
XX Amino acid sequence of human IRS-1 protein.
DE Phosphotyrosine interaction domain; PI domain; modulator; APP;
XX amyloid precursor protein; cytoplasmic domain; Alzheimer's disease;
KW neurodegenerative disease; dementia; IRS-1.
KW

XX Homo sapiens.
XX
XX WO9921995-A1.
XX
XX 06-MAY-1999.
XX
XX 23-OCT-1998; 98WO-US22523.
XX
XX 17-APR-1998; 98US-0062085.
PR 24-OCT-1997; 97US-0957660.
XX
XX (UYRQ) UNIV ROCKEFELLER.
XX

XX Buxbaum J, Greengard P, Sabo S;
PI WPI; 1999-326703/27.
XX
XX Cell line for screening agents that inhibit processing of amyloid
PT precursor protein
XX

PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0140354.
PR 18-JUN-1999; 99US-0140354.
PR 28-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.

PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 12-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149802.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 06-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.


```

PR 10-JAN-1991; 91US-0639404.
XX
XX (UYTE-) UNIV TENNESSEE.
XX
XX Brian DA, Kapke PA;
XX
XX WPI; 1993-143061/17.
XX
XX DR N-PSDB; AAQ40490.
XX
XX Porcine transmissible gastroenteritis virus DNA coding for
XX nucleocapsid and matrix proteins - used for diagnosis of
XX infection and in prodn. of vaccines
XX
XX Example 1; Columns 17-20; 13pp; English.
XX
XX Clone RG5 was the longest cDNA clone from the 3' end of the
XX porcine TGEV genome to be isolated by the inventors. The largest
XX ORF is predicted to encode a 382 amino acid protein of mol. wt.
XX 43,426. The only TGEV protein described to date that approaches
XX this size is the phosphorylated nucleocapsid protein (46-50 kD by
XX SDS-PAGE. See also AAQ40491.
XX
XX Sequence 382 AA;
SQ
Query Match 3.6%; Score 7; DB 14; Length 382;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 RKSRSKS 135
Db 342 RKSRSKS 348
|||||||

RESULT 32
ABE77059
ID ABB77059 standard; Protein; 458 AA.
XX
XX ABB77059;
XX
XX 08-OCT-2002 (first entry)
XX
XX Human protein sequence encoded by gene no:17: from cDNA clone HKAE128.
XX
XX Human; HKAE128; secreted protein; immunosuppressive; food preservative;
XX antithratic; antirheumatic; antiproliferative; cytostatic; cardiant;
XX vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
XX virucide; fungicide; ophthalmological; vulnery; gene therapy; ELISA;
XX radioimmuncassay; enzyme linked immunosorbent assay; autoimmune disease;
XX hyperproliferative disorder; cardiovascular disorder; angiogenesis;
XX cerebrovascular disorder; nervous system disorder; ocular disorder;
XX wound healing; food additive.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX
XX Misc-difference 403 /label= Unknown
XX
XX Misc-difference 455 /label= Unknown
XX
XX WO200222638-A1.
XX
XX 21-MAR-2002.
XX
XX 17-JAN-2001; 2001WO-US01386.
XX
XX 12-SEP-2000; 2000US-232104P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Komatsoulis GA, Baker KP, Birse CE, Soppet DR, Olsen HS;
XX Moore PA, Wei P, Ebner R, Duan DR, Shi Y, Choi GH, Fiscella M;
XX NI J;

10-JAN-1991; 258041/30.
XX
XX New nucleic acid molecules encoding 22 human secreted proteins for
XX diagnosing or treating e.g. autoimmune diseases, hyperproliferative
XX disorders, and cardiovascular disorders, and used as food additives or
XX preservatives -
XX
XX Disclosure; Page 49; 526pp; English.
XX
XX The sequence represents a protein sequence of the invention, encoded by
XX cDNA isolated from human clone ID HKAE128. The invention relates to novel
XX isolated nucleic acid molecules encoding 22 human secreted proteins. The
XX proteins of the invention have immunosuppressive, antithratic,
XX antirheumatic, antiproliferative, cytostatic, cardiant, vasotropic,
XX cerebroprotective, nootropic, neuroprotective, antibacterial, virucide,
XX fungicide, ophthalmological, and vulnerary activity. The polynucleotides
XX may have a use in gene therapy. The polynucleotides and polypeptides
XX encoded by them are used to prevent, treat or ameliorate a medical
XX condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs,
XX chickens or sheep. The polynucleotides and polypeptides are also used in
XX diagnosing a pathological condition or susceptibility to a pathological
XX condition. The antibodies to the proteins can also be used in alleviating
XX symptoms associated with the disorders and in diagnostic immunoassays
XX e.g. radioimmunoassays or enzyme linked immunosorbent assays (ELISA).
XX Disorders which are diagnosed or treated include autoimmune diseases,
XX hyperproliferative disorders, cardiovascular disorders, cerebrovascular
XX disorders, angiogenesis, nervous system disorders, infections caused by
XX bacteria, viruses and fungi and ocular disorders. The polypeptides can
XX also be used to aid wound healing and epithelial cell proliferation. The
XX polypeptides can also be used as a food additive or preservative.
XX
XX Sequence 458 AA;
SQ
Query Match 3.6%; Score 7; DB 23; Length 458;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 RKSRSKS 135
Db 58 RKSRSKS 64
|||||||

RESULT 33
ABE54651
ID ABB54651 standard; Protein; 628 AA.
XX
XX ABB54651;
XX
XX AC
XX
XX DT 16-MAY-2002 (first entry)
XX
XX Lactococcus lactis protein pabB.
XX
XX DE
XX
XX KW
XX
XX Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
XX
XX OS Lactococcus lactis IL1403.
XX
XX PN FR2807446-A1.
XX
XX PD 12-OCT-2001.
XX
XX PF 11-APR-2000; 2000FR-0004630.
XX
XX PR 11-APR-2000; 2000FR-0004630.
XX
XX (INRG ) INRA INST NAT RECH AGRONOMIQUE.
XX
XX Bolotine A, Sorokine A, Renault P, Ehrlich SD;
XX
XX WPI; 2002-043418/06.
XX
XX New nucleotide sequence useful in the identification of Lactococcus
XX lactis and related species -
PT

```

CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematous, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 XX
 SQ Sequence 334 AA;

Query Match 3.6%; Score 7; DB 21; Length 334;
 Best Local Similarity 100.0%; Pred. No. 59;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 RKSRKS 135
 |||||
 Db 42 RKSRKS 48

RESULT 29
 AAB51364
 ID AAB51364 standard; Protein; 376 AA.

XX AAB51364;

XX 10-APR-2001 (first entry)

XX Feline infectious peritonitis virus nucleocapsid protein SEQ ID NO:2.

XX Feline infectious peritonitis virus; FIP; FIPV; nucleocapsid protein;
 KW vaccine; coronavirus infectious disease.

XX Feline infectious peritonitis virus.

XX JP2000302692-A.

XX 31-OCT-2000.

XX 21-APR-1999; 99JP-0114115.

XX 21-APR-1999; 99JP-0114115.

XX (KYOR-) KYORITSU SHOJI KK.

XX WPI; 2001-141402/15.

XX N-PSDB; AAF26980.

XX A new DNA vaccine against coronavirus infectious disease

XX Example; Page 20; 29pp; Japanese.

XX The present invention describes a DNA vaccine against a coronavirus
 CC infectious disease which comprises a DNA encoding the nucleocapsid
 CC protein (I) in feline infectious peritonitis (FIP) virus and an
 CC expression vector which can express (I) in an applied animal and
 CC physiological saline water. The DNA vaccine can be used for the
 CC prevention and the treatment of a coronavirus infectious disease such
 CC as FIP virus (FIPV). The present sequence represents the FIPV
 CC nucleocapsid protein.

XX Sequence 376 AA;

Query Match 3.6%; Score 7; DB 22; Length 376;
 Best Local Similarity 100.0%; Pred. No. 66;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 RKSRKS 135
 |||||
 Db 337 RKSRKS 343

RESULT 30
 AAP82001

ID AAP82001 standard; protein; 380 AA.

XX AAP82001;

XX 14-DEC-1990 (first entry)

XX Sequence of phenylalanine dehydrogenase (PADH) encoded by gene isolated
 DE from *Bacillus sphaericus* SCRC-R79a (FERM BP-1013).

XX Enzyme.

XX *Bacillus sphaericus* SCRC-R79a (FERM BP-1013).

XX EP256514-A.

XX 24-FEB-1988.

XX 12-AUG-1987; 87EP-0111696.

XX 27-NOV-1986; 86JP-0280654.

XX 02-AUG-1986; 86JP-0187852.

XX (SAGA) SAGAMI CHEM RES CENTRE.

XX Hibino Y, Asano Y, Okazaki N, Numao N;

XX WPI; 1988-051314/08.

XX N-PSDB; AAN82009.

XX Isolated gene for bacterial phenylalanine dehydrogenase -
 PT with derived plasmids and transformant(s), in phenylalanine
 PT prodn. from phenyl-pyruvic acid

XX Disclosure; 3-3; 50pp; English.

XX The PADH gene (AAN82009) is used to transform microorganisms, esp.
 CC *E.coli*. Culturing the transformed organisms will produce PADH which is
 CC used to convert phenylpyruvic acid (PPA, or its salt), and NH₄ ions (in
 CC presence of a reducing agent) to L-phenylalanine.

XX Sequence 380 AA;

Query Match 3.6%; Score 7; DB 9; Length 380;
 Best Local Similarity 100.0%; Pred. No. 66;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 IVKSDDI 32

Db 235 IVKSDDI 241

RESULT 31

AAR39224

ID AAR39224 standard; Protein; 382 AA.

XX AAR39224;

XX 09-AUG-1993 (first entry)

XX Nucleocapsid protein deduced from TGEV clone FG5.

XX Transmissible gastroenteritis virus; pig; vaccine; N protein.

XX Transmissible gastroenteritis Virus.

XX US5202430-A.

XX 13-APR-1993.

XX 23-APR-1987; 87US-0041704.

XX 23-APR-1987; 87US-0041704.

XX 16-JAN-1990; 90US-0465570.

Best Local Similarity 100.0%; Pred. NO. 56;		Best Local Similarity 100.0%; Pred. NO. 57;	
Matches 7; Conservative 0; Mismatches 0;		Matches 7; Conservative 0; Mismatches 0;	
0; Indels 0; Gaps 0;		0; Indels 0; Gaps 0;	
QY	159 KRKDKR 165	QY	89 RISRLD 95
Db	236 KRKDKR 242	Db	189 RISRLD 195
RESULT 27		RESULT 28	
AAU54873		AAB41766	
ID	AAU54873 standard; Protein; 324 AA.	ID	AAB41766 standard; Protein; 334 AA.
XX	AAU54873;	XX	AAB41766;
AC		AC	
XX		XX	
DT	27-FEB-2002 (first entry)	DT	08-FEB-2001 (first entry)
XX		XX	
DE	Propionibacterium acnes immunogenic protein #15769.	DE	Human ORFX ORF1530 polypeptide sequence SEQ ID NO:3060.
XX		XX	
KW	SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;	KW	Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW	uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;	KW	vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW	inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;	KW	anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
XX	dermatological; osteopathic; neuroprotectant.	KW	immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
XX		KW	hypotensive; dermatological; immunosuppressive; antiinflammatory;
OS	Propionibacterium acnes.	KW	antiviral; antibacterial; antifungal; antirheumatic; antichyroid;
XX		KW	antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
PN	WO200181581-A2.	KW	neurodegenerative disorder; osteoarthritis; graft vs host disease;
XX		KW	cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
PD		KW	cholesterol ester storage; systemic lupus erythematosus; infection;
XX		KW	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
PF	20-APR-2001; 2001WO-US12865.	KW	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
XX		KW	bone damage; cartilage damage; antiinflammatory disease; coagulation;
PR	21-APR-2000; 2000US-199047P.	KW	thrombosis; contraceptive.
PR	02-JUN-2000; 2000US-208841P.	XX	Homo sapiens.
PR	07-JUL-2000; 2000US-216747P.	XX	
XX		XX	WO200058473-A2.
PA	(CORI-) CORIXA CORP.	XX	
XX		XX	05-OCT-2000.
PI	Skelky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;	XX	
PI	L'malsonneuve J, Zhang Y, Jen S, Carter D;	XX	
XX		XX	31-MAR-2000; 2000WO-US08621.
DR	WPI; 2001-616774/71.	XX	
DR	N-PSDB; AAS59566.	PR	31-MAR-1999; 99US-0127607.
XX		PR	02-APR-1999; 99US-0127636.
XX		PR	05-APR-1999; 99US-0127728.
PT	Propionibacterium acnes polypeptides and nucleic acids useful for	PR	30-MAR-2000; 2000US-0540763.
PT	vaccinating against and diagnosing infections, especially useful for	XX	
PT	treating acne vulgaris .	XX	(CURA-) CURAGEN CORP.
XX		XX	
PS	Example 1; SEQ ID No 16068; 1069pp; English.	XX	Shimkets RA, Leach M;
XX		PI	
CC	Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic	XX	WPI; 2000-602362/57.
CC	polypeptides. The proteins and their associated DNA sequences are used in	DR	N-PSDB; AAC75975.
CC	the treatment, prevention and diagnosis of medical conditions caused by	XX	
CC	P. acnes. The disorders include SAPHO syndrome (synovitis, acne,	XX	
CC	pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.	PT	Novel nucleic acids and peptides derived from open reading frame X,
CC	P. acnes is also involved in infections of bone, joints and the central	PT	useful for treating e.g. cancers, proliferative disorders,
CC	nervous system, however it is particularly involved in the inflammatory	PT	neurodegenerative disorders and cardiovascular disease .
CC	lesions associated with acne vulgaris. A method for detecting the	XX	
CC	presence or absence of P. acnes in a patient comprises contacting a	PS	Claim 11; Page 2282-2283; 5507pp; English.
CC	sample with a binding agent that binds to the proteins of the invention	XX	
CC	and determining the amount of bound protein in the sample. The	XX	AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC	polypeptides may be used as antigens in the production of antibodies	CC	which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC	specific for P. acnes proteins. These antibodies can be used to	CC	sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC	downregulate expression and activity of P. acnes polypeptides and	CC	antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC	therefore treat P. acnes infections. The antibodies may also be used as	CC	osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC	diagnostic agents for determining P. acnes presence, for example, by	CC	immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC	enzyme linked immunosorbent assay (ELISA).	CC	antidiabetic; hypotensive; dermatological; immunosuppressive;
CC	Note: The sequence data for this patent did not form part of the printed	CC	antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC	specification, but was obtained in electronic format directly from WIPO	CC	antichyroid; and antianaemic. The sequences can be used for determining
CC	at ftp.wipo.int/pub/published_pct_sequences.	CC	the presence of or predisposition to, or preventing or treating
XX		CC	pathological conditions associated with an ORFX-associated disorder. The
XX		CC	nucleic acids can be used to express ORFX proteins in gene therapy
XX		CC	vectors. The proteins and nucleic acids may be used to treat cancers,
XX		CC	proliferative disorders, neurodegenerative disorders, osteoarthritis,
SQ	Sequence 324 AA;		
Query Match 3.6%; Score 7; DB 22; Length 324;			

```

Db      275 SKSQSSK 281
|||||||
RESULT 25
AAW73507
ID      AAW73507 standard; Protein; 317 AA.
XX
AC      AAW73507;
XX
DT      01-MAR-1999 (first entry)
XX
DE      Human ATG-1709 protein.
XX
KW      Human; ATG-1709 protein; secreted ligand; 7-Transmembrane receptor;
KW      heart disease; hypertension; cardiovascular disease; kidney disease;
KW      obesity; insulin resistance; diabetes; Central Nervous System disorder;
KW      therapy; SFRP-1.
XX
OS      Homo sapiens.
XX
PN      EP879885-A1.
XX
PD      25-NOV-1998.
XX
PF      16-JAN-1998; 98EP-0300313.
XX
PR      08-AUG-1997; 97US-0907808.
PR      23-MAY-1997; 97US-0047691.
XX
PA      (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI      Hu E, Zhu Y;
XX
WPI: 1998-596877/51.
DR      N-PSDB; AAV08946.
XX
New human secreted protein ATG-1709 polypeptide and polynucleotide -
PT      useful as diagnostic reagents and for diagnosing, prevention and
PT      treatment of Central Nervous System diseases and diabetes
XX
Claim 11; Page 6; 28pp; English.
XX
This sequence represents the human ATG-1709 protein of the invention.
CC      ATG-1709 is related to human secreted ligands for 7-Transmembrane
CC      receptors and similar to murine sfrp-1. ATG-1709 polypeptides and
CC      polynucleotides are useful for diagnosing susceptibility to diseases by
CC      detecting mutations in the ATG-1709 gene using probes containing the
CC      ATG-1709 nucleotide sequence, and can diagnose diseases associated with
CC      ATG-1709 imbalance by determining ATG-1709 polypeptide expression levels.
CC      ATG-1709 polypeptides can be used to screen for agonists and antagonists
CC      which bind the ATG-1709 polypeptide. These can be used in treatment to
CC      activate or inhibit ATG-1709 activity, in addition to direct
CC      administration of antisense sequences to prevent expression, or ATG-1709
CC      polypeptides to treat conditions associated with a lack of ATG-1709.
CC      Gene therapy may also be used to affect endogenous ATG-1709 expression.
CC      ATG-1709 antibodies are useful for inducing an immune response to
CC      immunise and prevent diseases, and for isolating ATG-1709 clones or
CC      purifying the polypeptides by affinity chromatography. ATG-1709
CC      polypeptides can be administered directly or as a vaccine to inoculate
CC      against disease. Diseases diagnosed, prevented or treated include:
CC      heart disease; hypertension; cardiovascular diseases; kidney diseases;
CC      obesity; insulin resistance; diabetes and Central Nervous System (CNS)
CC      diseases. The ATG-1709 polypeptide is also useful for mapping the gene to
CC      a chromosome, allowing gene inheritance to be studied through linkage
CC      analysis.
XX
SQ      Sequence 317 AA;
Query Match 3.6%; Score 7; DB 19; Length 317;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      159 KRKDKR 165
|||||||
Db      236 KRKDKR 242
|||||||
RESULT 26
AAW37816
ID      AAW37816 standard; Protein; 317 AA.
XX
AC      AAW37816;
XX
DT      28-AUG-1998 (first entry)
XX
DE      Human secreted apoptosis-related protein hSARP3.
XX
KW      Secreted apoptosis-related protein; SARP; hSARP3; human;
KW      prostate cancer; breast cancer; diagnosis; gene therapy.
XX
OS      Homo sapiens.
XX
FH      Key Location/Qualifiers
FT      Peptide 1..28
FT      Protein /label= Sig_peptide
FT      29..317
FT      /label= Mat_protein
XX
PN      WO9813493-A2.
XX
PD      02-APR-1998.
XX
PF      24-SEP-1997; 97WO-US17154.
XX
PR      11-OCT-1996; 96US-0028363.
PR      24-SEP-1996; 96US-0026603.
XX
PA      (LXRB-) LXR BIOTECHNOLOGY INC.
XX
PI      Melkonyan H, Umansky S;
XX
WPI: 1998-230704/20.
DR      N-PSDB; AAV19114.
XX
New secreted apoptosis-related proteins - useful for modulating
PT      apoptosis, particularly for treatment of prostatic or breast cancer,
PT      also for diagnosis and monitoring of disease
XX
Claim 1; Page 56-58; 101pp; English.
XX
This polypeptide comprises human secreted apoptosis-related
CC      protein hSARP3 that modulates apoptosis through cell-cell or
CC      cell-extracellular matrix signalling. Its amino acid sequence was
CC      deduced from a clone (see AAV19114) obtained from a human pancreas
CC      cDNA library. Murine hSARP1 (see AAW37814), human hSARP1 (see
CC      AAW37816) and hSARP2 (see AAW37817) proteins are also claimed. hSARP3
CC      is expressed predominantly in pancreas. SARP polypeptides can be
CC      obtained from recombinant host cells. Antibodies specific for SARP
CC      polypeptides can be used in immunoassays for detecting levels of
CC      expression of SARP, particularly for diagnosis or monitoring of
CC      diseases associated with SARP expression. Specifically, they are
CC      used to detect cancer of the prostate or breast (by detecting
CC      hSARP1 and 2, respectively). SARP polypeptides and nucleic acids
CC      can also be used to treat these cancers, or more generally
CC      apoptosis-related disease (e.g. infection with HIV or reperfusion
CC      injury), also (not claimed) to prevent apoptosis in cultured cells,
CC      to improve preservation of organs for transplantation, for in situ
CC      preservation for by-pass operations and to treat dermatological
CC      disorders. SARP polypeptides can also be used to identify agents,
CC      potentially useful therapeutically, that modulate the effects of
CC      SARP on Wnt-frizzled protein interaction.
XX
SQ      Sequence 317 AA;
Query Match 3.6%; Score 7; DB 19; Length 317;

```

Db 260 YFVLENP 266

RESULT 23
AAM41951
ID AAM41951 standard; Protein; 283 AA.
XX AC AAM41951;
XX DT 22-OCT-2001 (first entry)
XX DE Human polypeptide SEQ ID NO 6882.
XX KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX OS Homo sapiens.
XX PN WO200153312-A1.
XX PD 26-JUL-2001.
XX PF 26-DEC-2000; 2000WO-US34263.
XX PR 21-JAN-2000; 2000US-0488725.
XX PR 25-APR-2000; 2000US-0552317.
XX PR 09-JUL-2000; 2000US-0598042.
XX PR 19-JUL-2000; 2000US-0620312.
XX PR 03-AUG-2000; 2000US-0653450.
XX PR 14-SEP-2000; 2000US-0662191.
XX PR 19-OCT-2000; 2000US-0693036.
XX PR 29-NOV-2000; 2000US-0727344.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Dermanac RT;
XX WPI: 2001-442253/47.
XX N-PSDB; AAI61107.
XX Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
XX Example 2; SEQ ID NO 6882; 10078pp; English.
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
XX the encoded polypeptides (AAM38642-AAM42213) with nootropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: Immune system suppression,
XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukaemias and
XX C.N.S disorders.
XX Note: The sequence data for this patent did not form part of the printed
XX specification.
SQ Sequence 283 AA;
Query Match 3.6%; Score 7; DB 22; Length 283;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 YFVLENP 113
Db 263 YFVLENP 269
|||||
RESULT 24
AAY41211
ID AAY41211 standard; Protein; 284 AA.
XX AC AAY41211;
XX DT 31-JAN-2000 (first entry)
XX DE H. influenzae Yigt protein.
XX KW Membrane targeting protein; translocation protein; Escherichia coli;
KW Sec-independent pathway; protein transport; twin-arginine signal peptide;
KW mtABC operon; Mta protein; Yigt protein.
XX OS Haemophilus influenzae.
XX FH Key Location/Qualifiers
XX FT Misc-difference 110
XX FT /note= "unspecified"
XX PN WO9951753-A1.
XX PD 14-OCT-1999.
XX PF 29-MAR-1999; 99WO-CA00272.
XX PR 01-APR-1998; 98US-0053197.
XX PR 28-MAY-1998; 98US-0085761.
XX PA (UYAL-) UNIV ALBERTA.
XX PI Weiner JH, Turner RJ;
XX WPI: 1999-633740/54.
XX New recombinant membrane targeting and translocation proteins from
XX Escherichia coli, used to produce soluble polypeptides normally
XX produced in insoluble form -
XX Disclosure: Fig 6; 11pp; English.
XX The invention relates to recombinant membrane targeting and translocation
XX proteins from Escherichia coli. The recombinant polypeptides are involved
XX in the Sec-independent pathway for transporting proteins with a twin-
XX arginine signal peptide to the periplasm or extracellular medium and to
XX the cell membrane. They transport fully folded and co-factor containing
XX proteins. The polypeptides are used to transport such proteins,
XX specifically to produce soluble forms of polypeptides that are normally
XX produced in insoluble form. They may also be used to raise specific
XX antibodies. Nucleic acid sequences that encode the polypeptides are used
XX for production of recombinant proteins and their fragments are used as
XX probes to detect or isolate related genomic or cDNA sequences (these have
XX been found in many other bacteria, yeast, plants, nematodes and humans).
XX producing normally insoluble proteins in soluble form eliminates the need
XX for resolubilization with strong denaturants and facilitates recovery of
XX functional proteins (which have been properly folded by cytoplasmic
XX enzymes before translocation). The present sequence represents the
XX H. influenzae Yigt protein which has homology to the E. coli Mta
XX protein.
SQ Sequence 284 AA;
Query Match 3.6%; Score 7; DB 20; Length 284;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 SKSQSSK 139

RESULT 21

AAU30761
ID AAU30761 standard; Protein; 266 AA.

XX AC AAU30761;

XX DT 18-DEC-2001 (first entry)

XX DE Novel human secreted protein #1252.

XX KW Human; vaccination; gene therapy; nutritional supplement;

XX KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;

XX KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.

XX OS Homo sapiens.

XX PN WO200179449-A2.

XX PD 25-OCT-2001.

XX PF 16-APR-2001; 2001WO-US08656.

XX PR 18-APR-2000; 2000US-0552929.

XX PR 26-JAN-2001; 2001US-0770160.

XX XX (HYSE-) HYSEQ INC.

XX XX Tang YT, Liu C, Drmanac RT;

XX XX WPI; 2001-611725/70.

XX XX Nucleic acids encoding a range of human polypeptides, useful in genetic
XX PT vaccination, testing and therapy -

XX PS Claim 20; Page 341; 765pp; English.

XX CC The invention relates to novel human secreted polypeptides. The
XX CC polypeptides and antibodies to the polypeptides are useful for
XX CC determining the presence of or predisposition to a disease associated
XX CC with altered levels of polypeptide. The polypeptides are also useful for
XX CC identifying agents (agonists and antagonists) that bind to them. Cells
XX CC expressing the proteins are useful for identifying a therapeutic agent
XX CC for use in treatment of a pathology related to aberrant expression or
XX CC physiological interactions of the polypeptide. Vectors comprising
XX CC the nucleic acids encoding the polypeptides and cells genetically
XX CC engineered to express them are also useful for producing the proteins.
XX CC The proteins are useful in genetic vaccination, testing and
XX CC therapy, and can be used as nutritional supplements. They may be used to
XX CC increase stem cell proliferation; to regulate haematopoiesis; and in
XX CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
XX CC immune suppression and/or stimulation; as anti-inflammatory agents; and
XX CC in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid
XX CC sequences of novel human secreted proteins of the invention.

XX SQ Sequence 266 AA;

Query Match 3.6%; Score 7; DB 22; Length 266;

Best Local Similarity 100.0%; Pred. No. 47;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 124 RPHKRRK 130

|||||

Db 157 RPHKRRK 163

RESULT 22

AAB42139
ID RAB42139 standard; Protein; 280 AA.

XX AC AAB42139;

XX DT 08-FEB-2001 (first entry)

XX

Human ORFX ORF1903 polypeptide sequence SEQ ID NO:3806.

DE XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.

XX OS Homo sapiens.

XX PN WO200058473-A2.

XX PD 05-OCT-2000.

XX PF 31-MAR-2000; 2000WO-US08621.

XX XX 31-MAR-1999; 99US-0127607.

XX PR 02-APR-1999; 99US-0127636.

XX PR 05-APR-1999; 99US-0127728.

XX PR 30-MAR-2000; 2000US-0540763.

XX PA (CURA-) CURAGEN CORP.

XX PI Shinkets RA, Leach M;

XX XX WPI; 2000-602362/57.

XX DR N-PSDB; AAC76348.

XX DR Novel nucleic acids and peptides derived from open reading frame X,
XX PT useful for treating e.g. cancers, proliferative disorders,
XX PT neurodegenerative disorders and cardiovascular disease -

PS Claim 11; Page 2956; 5507pp; English.

XX CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
XX CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
XX CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
XX CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
XX CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
XX CC antidiabetic; hypotensive; dermatological; immunosuppressive;
XX CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
XX CC antithyroid; and antianaemic. The sequences can be used for determining
XX CC the presence of or predisposition to, or preventing or treating
XX CC pathological conditions associated with an ORFX-associated disorder. The
XX CC nucleic acids can be used to express ORFX proteins in gene therapy
XX CC vectors. The proteins and nucleic acids may be used to treat cancers,
XX CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
XX CC graft vs host disease, cardiovascular disease, diabetes mellitus,
XX CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
XX CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
XX CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
XX CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
XX CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
XX CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX SQ Sequence 280 AA;

Query Match 3.6%; Score 7; DB 21; Length 280;

Best Local Similarity 100.0%; Pred. No. 50;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 YFVLENP 113

|||||

PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.

PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 3.6%; Score 7; DB 21; Length 256;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 KEVTPEM 9
| | | | |
Db 117 KEVTPEM 123

RESULT 19	Matches	7:	Conservative	0:	Mismatches	0:	Indels	0:	Gaps	0:
AAB36598	QY	107	YFVLENP 113							
ID AAB36598 standard; Protein; 254 AA.	Db	234	YFVLENP 240							
XX AAB36598;										
XX										
DT 09-MAR-2001 (first entry)										
XX										
DE Human FLEXHT-20 protein sequence SEQ ID NO:20.										
XX										
KW Human; FLEXHT: full-length molecules expressed in human tissue;										
KW diagnosis; gene expression; genetic linkage; genetic variability;										
KW antianemic; anticonvulsant; antiarteriosclerotic; immunomodulatory;										
KW cytosolic; antiasthmatic; antiinflammatory; hepatotropic; antidiabetic;										
KW anti-gout; antithyroid; neuroprotective; antiarthritic; osteopathic;										
KW antipsoriatic; antirheumatic; antiulcer; gene therapy; anaemia; gout;										
KW epilepsy; arteriosclerosis; atherosclerosis; developmental disorder;										
KW cancer; immunological disorder; asthma; bronchitis; cirrhosis;										
KW Crohn's disease; diabetes mellitus; Grave's disease; multiple sclerosis;										
KW osteoarthritis; pancreatitis; rheumatoid arthritis; psoriasis;										
KW ulcerative colitis.										
XX										
OS Homo sapiens.										
XX										
PN WO200070047-A2.										
XX										
PD 23-NOV-2000.										
XX										
PF 12-MAY-2000; 2000WO-US13299.										
XX										
PR 14-MAY-1999; 99US-0311894.										
PR 14-MAY-1999; 99US-0311937.										
PR 14-MAY-1999; 99US-0311940.										
XX										
PA (INCY-) INCYTE GENOMICS INC.										
XX										
PI Yue H, Tang YT, Lal P, Reddy R, Batra S, Baughn MR, Yang J;										
PI Azimzai Y, Lu DAM, Au-Young J, Shih LL;										
XX										
PI WPI: 2001-016234/02.										
DR N-PSDB; AAC88089.										
XX										
PT Human FLEXHT protein and DNA sequences, useful for treating										
PT immunological disorders, developmental disorders, and cancers -										
XX										
PS Claim 1; Page 113; 168pp; English.										
XX										
CC AAC88070 to AAC88124 encode the 55 FLEXHT (full-length molecules										
CC expressed in human tissues) proteins given in AAB36579 to AAB36633. The										
CC present invention describes an isolated polypeptide (A) comprising an										
CC amino acid sequence selected from one of 55 amino acid sequences 42-876										
CC residues in length, corresponding to FLEXHT-1 to FLEXHT-55, a 90 %										
CC identical sequence, and a biologically active or immunogenic fragment of										
CC the sequence. The FLEXHT proteins can have antianemic, anticonvulsant,										
CC antiarteriosclerotic, immunomodulatory, cytosolic, antiasthmatic,										
CC antiinflammatory, hepatotropic, antidiabetic, anti-gout, antithyroid,										
CC neuroprotective, antiarthritic, osteopathic, antipsoriatic, antiulcer										
CC and antirheumatic activities, and can be used in gene therapy. The										
CC polynucleotide sequences can be used to express the protein sequences.										
CC Pharmaceutical compositions comprising FLEXHT can be used to treat										
CC diseases or conditions associated with altered expression of functional										
CC FLEXHT. The proteins and polynucleotides can be used to diagnose and										
CC treat disorders including anaemia, epilepsy, arteriosclerosis,										
CC atherosclerosis, developmental disorders, cancers, and immunological										
CC disorders such as asthma, bronchitis, cirrhosis, Crohn's disease,										
CC diabetes mellitus, gout, Grave's disease, multiple sclerosis,										
CC osteoarthritis, pancreatitis, psoriasis, rheumatoid arthritis, and										
CC ulcerative colitis.										
XX										
SQ Sequence 254 AA;										
Query Match	3.6%;	Score 7;	DB 22;	Length 254;						
Best Local Similarity	100.0%;	Pred. No. 45;								

PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140595.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142520.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.

PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 05-OCT-1999; 99US-0157117.
PR 06-OCT-1999; 99US-0157753.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 3.6%; Score 7; DB 21; Length 248;
Best Local Similarity 100.0%; Pred No. 44;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 3 KEVTPM 9
|||||
Db 109 KEVTPM 115

PR	06-AUG-1999;	99US-0147303.	Matches	7;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
PR	06-AUG-1999;	99US-0147416.	Qy	3	KEVTPEM 9							
PR	09-AUG-1999;	99US-0147493.										
PR	09-AUG-1999;	99US-0147935.	Db	101	KEVTPEM 107							
PR	10-AUG-1999;	99US-0148171.										
PR	11-AUG-1999;	99US-0148319.										
PR	12-AUG-1999;	99US-0148341.										
PR	13-AUG-1999;	99US-0148565.										
PR	13-AUG-1999;	99US-0148684.										
PR	16-AUG-1999;	99US-0149368.										
PR	17-AUG-1999;	99US-0149175.										
PR	18-AUG-1999;	99US-0149426.										
PR	20-AUG-1999;	99US-0149722.										
PR	20-AUG-1999;	99US-0149723.										
PR	20-AUG-1999;	99US-0149929.										
PR	23-AUG-1999;	99US-0149902.										
PR	23-AUG-1999;	99US-0149930.										
PR	25-AUG-1999;	99US-0150566.										
PR	26-AUG-1999;	99US-0150884.										
PR	27-AUG-1999;	99US-0151065.										
PR	27-AUG-1999;	99US-0151066.										
PR	27-AUG-1999;	99US-0151080.										
PR	30-AUG-1999;	99US-0151303.										
PR	31-AUG-1999;	99US-0151436.										
PR	01-SEP-1999;	99US-0151930.										
PR	07-SEP-1999;	99US-0152363.										
PR	10-SEP-1999;	99US-0153070.										
PR	13-SEP-1999;	99US-0153758.										
PR	15-SEP-1999;	99US-0154018.										
PR	16-SEP-1999;	99US-0154039.										
PR	20-SEP-1999;	99US-0154779.										
PR	22-SEP-1999;	99US-0155139.										
PR	23-SEP-1999;	99US-0155486.										
PR	24-SEP-1999;	99US-0155659.										
PR	28-SEP-1999;	99US-0156458.										
PR	29-SEP-1999;	99US-0156596.										
PR	04-OCT-1999;	99US-0157117.										
PR	05-OCT-1999;	99US-0157753.										
PR	06-OCT-1999;	99US-0157865.										
PR	07-OCT-1999;	99US-0158029.										
PR	08-OCT-1999;	99US-0158232.										
PR	12-OCT-1999;	99US-0158369.										
PR	13-OCT-1999;	99US-0159293.										
PR	13-OCT-1999;	99US-0159294.										
PR	13-OCT-1999;	99US-0159295.										
PR	14-OCT-1999;	99US-0159329.										
PR	14-OCT-1999;	99US-0159330.										
PR	14-OCT-1999;	99US-0159331.										
PR	14-OCT-1999;	99US-0159637.										
PR	14-OCT-1999;	99US-0159638.										
PR	18-OCT-1999;	99US-0159584.										
PR	21-OCT-1999;	99US-0160741.										
PR	21-OCT-1999;	99US-0160767.										
PR	21-OCT-1999;	99US-0160768.										
PR	21-OCT-1999;	99US-0160770.										
PR	21-OCT-1999;	99US-0160814.										
PR	21-OCT-1999;	99US-0160815.										
PR	22-OCT-1999;	99US-0160980.										
PR	22-OCT-1999;	99US-0160981.										
PR	22-OCT-1999;	99US-0160989.										
PR	25-OCT-1999;	99US-0161404.										
PR	25-OCT-1999;	99US-0161405.										
PR	25-OCT-1999;	99US-0161406.										
PR	26-OCT-1999;	99US-0161359.										
PR	26-OCT-1999;	99US-0161360.										
PR	26-OCT-1999;	99US-0161361.										
PR	28-OCT-1999;	99US-0161920.										
PR	28-OCT-1999;	99US-0161992.										
PR	28-OCT-1999;	99US-0161993.										
PR	29-OCT-1999;	99US-0162142.										

Query Match 3.6%; Score 7; DB 21; Length 240;
Best Local Similarity 100.0%; Pred. No. 43;

CC Note: The sequence data for this patent did not form part of the printed
CC specification.

XX
SQ Sequence 235 AA;

Query Match 3.6%; Score 7; DB 22; Length 235;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 YFVLENP 113
|||||||
Db 215 YFVLENP 221

RESULT 17

AAG09871
ID AAG09871 standard; Protein; 240 AA.

XX

AC AAG09871;

XX DT 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 7970.

XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX OS Arabidopsis thaliana.
XX EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

XX PR 05-MAR-1999; 99US-0123180.

XX PR 09-MAR-1999; 99US-0123548.

XX PR 23-MAR-1999; 99US-0125788.

XX PR 25-MAR-1999; 99US-0126264.

XX PR 29-MAR-1999; 99US-0126785.

XX PR 01-APR-1999; 99US-0127462.

XX PR 06-APR-1999; 99US-0128234.

XX PR 08-APR-1999; 99US-0128714.

XX PR 16-APR-1999; 99US-0129845.

XX PR 19-APR-1999; 99US-0130077.

XX PR 21-APR-1999; 99US-0130449.

XX PR 23-APR-1999; 99US-0130510.

XX PR 28-APR-1999; 99US-0131449.

XX PR 30-APR-1999; 99US-0132048.

XX PR 04-MAY-1999; 99US-0132407.

XX PR 05-MAY-1999; 99US-0132485.

XX PR 06-MAY-1999; 99US-0132486.

XX PR 06-MAY-1999; 99US-0132487.

XX PR 07-MAY-1999; 99US-0132863.

XX PR 11-MAY-1999; 99US-0134256.

XX PR 14-MAY-1999; 99US-0134218.

XX PR 14-MAY-1999; 99US-0134219.

XX PR 14-MAY-1999; 99US-0134221.

XX PR 14-MAY-1999; 99US-0134370.

XX PR 18-MAY-1999; 99US-0134768.

XX PR 19-MAY-1999; 99US-0134941.

XX PR 20-MAY-1999; 99US-0135124.

XX PR 21-MAY-1999; 99US-0135353.

XX PR 24-MAY-1999; 99US-0135629.

XX PR 25-MAY-1999; 99US-0136021.

XX PR 27-MAY-1999; 99US-0136392.

XX PR 28-MAY-1999; 99US-0136782.

XX PR 01-JUN-1999; 99US-0137222.

XX PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.

CC useful for treating chronic neurodegenerative diseases such as
 CC Alzheimer's disease and Parkinsonian's disease, autoimmune diseases,
 CC diabetes mellitus, congenital muscular dystrophy with mitochondrial
 CC structural abnormalities, fatal infantile myopathy, mitochondrial
 CC encephalopathy, lactic acidosis, stroke, mitochondrial diabetes,
 CC deafness, neuropathy, progressive external ophthalmoplegia, optic
 CC atrophy, Leigh's syndrome, dystonia, stroke, schizophrenia, progressive
 CC joint disorders such as osteoarthritis, and hyperproliferative
 CC disorders such as cancer, tumour and psoriasis.
 CC

SQ Sequence 157 AA;

Query Match 3.6%; Score 7; DB 22; Length 157;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 53 FSEILK 59
 |||||
 Db 53 FSEILK 59

RESULT 15

AAW93976
 ID AAW93976 standard; protein; 159 AA.

XX AAW93976;

AC 30-JUN-1999 (first entry)

DE Human IRS-1 and IRS-2 binding inhibitor peptide 4.

XX Insulin receptor substrate; IRS-1; IRS-2; inhibitor; disease; obesity;
 KW insulin resistance; 14-3-3 protein; inhibitor; screening; diabetes;
 KW diabetic retinopathy; diabetic neuropathy; glucose tolerance;
 KW diabetic nephropathy; hyperinsulinaemia; hyperlipemia; arteriosclerosis;
 KW hypertension; ischaemic heart disease; ischaemic brain disease;
 KW peripheral embolism.

OS Homo sapiens.

XX WO9916462-A1.

PN 08-APR-1999.

XX 25-SEP-1998; 98WO-JP04293.

XX 29-SEP-1997; 97JP-0263719.

XX (DAUC) DAIICHI PHARM CO LTD.

XX Asano T, Kanda A, Kubo H, Yazaki Y;

XX WPI: 1999-254929/21.

XX Treatment of insulin resistance using insulin receptor substrate
 PT binding inhibitor

XX Disclosure; Page 24-25; 30pp; Japanese.

CC This invention describes a method for the treatment of diseases involving
 CC insulin resistance using a substance which inhibits the binding of
 CC insulin receptor substrate to 14-3-3 protein, identified by screening
 CC potential inhibitors for their ability to inhibit this binding. The
 CC composition described in the invention for the treatment of diseases
 CC involving insulin resistance contains as active component, an inhibitor
 CC of the binding of all or part of insulin receptor substrate 1 or 2
 CC (IRS-1, IRS-2) to all or part of 14-3-3 protein. Disorders involving
 CC insulin resistance include diabetes, diabetic retinopathy, diabetic
 CC neuropathy, impaired glucose tolerance, diabetic nephropathy,
 CC hyperinsulinaemia, hyperlipemia, arteriosclerosis, hypertension,
 CC obesity, ischaemic heart disease, ischaemic brain disease, and peripheral
 CC embolism.
 CC

SQ Sequence 159 AA;

Query Match 3.6%; Score 7; DB 20; Length 159;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 132 RSKSQSS 138
 |||||

Db 108 RSKSQSS 114

RESULT 16

AAW40165
 ID AAW40165 standard; Protein; 235 AA.

XX AAW40165;

AC 22-OCT-2001 (first entry)

DE Human polypeptide SEQ ID NO 3310.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.

XX Homo sapiens.

XX WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.

XX 25-APR-2000; 2000US-052317.

XX 09-JUL-2000; 2000US-0598042.

XX 19-JUL-2000; 2000US-0620312.

XX 03-AUG-2000; 2000US-0653450.

XX 14-SEP-2000; 2000US-0662191.

XX 19-OCT-2000; 2000US-0693036.

XX 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI: 2001-442253/47.

XX N-PSDB: AAI59321.

XX Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -

XX Example 5; SEQ ID NO 3310; 10078pp; English.

CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
 CC the encoded polypeptides (AAW38642-AAW42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Actin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.

CC S. epidermidis polypeptides (II) via the production of vectors
 CC containing them which are used to produce hosts cells which express the
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
 CC used to vaccinate subjects and to raise antibodies against the bacteria.
 CC The polypeptides may also be used to assay for other inhibitors of their
 CC activity and therefore identify compounds that may be used for the
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
 CC AAH5090 represent specifically claimed S. epidermidis genomic DNA
 CC polynucleotide sequences from the present invention. AAH55091 to
 CC AAH55098 represent oligonucleotide sequences and primers which are used
 CC in the exemplification of the present invention.
 CC N.B. The present invention specifically claims all the polynucleotide
 CC sequences given in the sequence listing of the present specification,
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
 CC no sequences are present for SEQ ID NO:4455 to 4464.

XX SQ Sequence 129 AA;

Query Match 3.6%; Score 7; DB 22; Length 129;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 EYCNFGC 105
 Db 75 EYCNFGC 81
 |||||

RESULT 13
 ABP40049
 ID ABP40049 standard; Protein; 132 AA.

XX ABP40049;

XX 24-JUL-2002 (first entry)

DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4894.

XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
 KW antibacterial; gene therapy.

XX Staphylococcus epidermidis.

XX US6380370-B1.

XX 30-APR-2002.

XX 13-AUG-1998; 98US-0134001.

XX 14-AUG-1997; 97US-055779P.

PR 08-NOV-1997; 97US-064964P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Doucette-Stamm LA, Bush D;

XX WPI; 2002-381255/41.

DR N-PSDB; ABN92594.

XX Novel isolated nucleic acid encoding a Staphylococcus epidermidis
 PT polypeptide, useful for diagnosing and treating bacterial infections -
 PS Disclosure; SEQ ID 4894; 267pp; English.

XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
 CC given in ABP35124 to ABP37960. The S. epidermidis sequences have
 CC antibacterial activity and can be used in gene therapy. The sequences
 CC can also be used in the diagnosis and treatment of bacterial infections,
 CC particularly S. epidermidis infections. The sequences can be used to
 CC screen for compounds able to interfere with the S. epidermidis life
 CC cycle or inhibit S. epidermidis infection.
 CC N.B. The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from the
 CC USPTO web site.

XX SQ Sequence 132 AA;

Query Match 3.6%; Score 7; DB 23; Length 132;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 EYCNFGC 105
 Db 78 EYCNFGC 84
 |||||

RESULT 14
 AAB47229
 ID AAB47229 standard; Protein; 157 AA.

XX AAB47229;

XX 18-JUL-2001 (first entry)

XX Yeast Mfn.

XX Mitofusin; homolog; GTPase; mediation; mitochondrial fusion; stroke;
 KW post-meiotic fusion; mitochondria; spermatid; anti-insect; deafness;
 KW antifungal; fertility; mortality; arthropod pest; Leigh's syndrome;
 KW mitochondrial disease; muscle myopathy; neurodegenerative disorder;
 KW aging; Alzheimer's disease; Parkinsonian's disease; cancer; psoriasis;
 KW autoimmune disease; diabetes mellitus; congenital muscular dystrophy;
 KW fatal infantile myopathy; mitochondrial encephalopathy; optic atrophy;
 KW lactic acidosis; mitochondrial diabetes; neuropathy; osteoarthritis;
 KW progressive external ophthalmoplegia; dystonia; schizophrenia.

XX Saccharomyces cerevisiae.

XX WO200125274-A1.

XX 12-APR-2001.

XX 06-OCT-2000; 2000WO-US27871.

XX 06-OCT-1999; 99US-0413285.

XX (STRD) UNIV LELAND STANFORD JUNIOR.

XX Fuller MT, Hales KG, Santel AH;

XX WPI; 2001-300165/31.

XX New nucleic acid molecules encoding mitofusin polypeptides useful for
 PT identifying anti-insect, antifungal and therapeutic agents -

XX Disclosure; Page 83-84; 92pp; English.

XX The sequences given in AAB47221-30 show mitofusin proteins and their
 CC homologs. Mitofusins are large predicted GTPases with a predicted
 CC transmembrane domain, coiled-coil region and a C-terminal region showing
 CC a high isoelectric point and a predicted coiled-coil region. They are
 CC the first known protein mediators of mitochondrial fusion, e.g.
 CC mediating developmentally regulated post-meiotic fusion of mitochondria
 CC in Drosophila spermatids. Mitofusin polypeptides are useful for
 CC enhancing membrane fusion in a population of membrane-bounded entities,
 CC e.g., mitochondria. Mitofusin proteins are useful for screening
 CC anti-insect or antifungal agent. Modulators of mitochondrial fusion
 CC are useful for reducing fertility and increasing the mortality rate
 CC of arthropod pest, e.g., an insect (claimed). Mitofusin DNA is
 CC useful for generating genetically modified non-human animals or
 CC site-specific gene modifications in cell lines, for the study of
 CC mitofusin function or regulation, and to create animal models of
 CC diseases, including mitochondrial diseases, muscle myopathies,
 CC neurodegenerative disorders, and aging. Mitofusins, and the DNA
 CC encoding them, can be used in assays to identify therapeutic agents

XX WO9934015-A2.
 PN 08-JUL-1999.
 PD 22-DEC-1998; 98WO-GB03887.
 PF 27-FEB-1998; 98GB-0004050.
 PR 24-DEC-1997; 97GB-0027347.
 XX (IMCR) IMPERIAL CANCER RES TECHNOLOGY.
 PA Ganesan TS;
 XX WPI: 1999-405519/34.
 PI N-PSDB; AAX84493.
 DR Methods of diagnosis, prognosis and treatment of cancer
 XX Disclosure; Fig 6; 167pp; English.
 CC This sequence is the p90 ribosomal S6 kinase-3 (Rsk-3).
 CC The invention relates of diagnosis, prognosis and treatment of cancer
 CC related to the Rsk-3 gene. The methods are used for diagnosis, prognosis
 CC and treatment of cancer, especially ovarian or breast cancer or lymphoma.
 CC Nucleic acids that selectively hybridise to the Rsk-3 gene or cDNA, or a
 CC mutant Rsk-3 allele, or a molecule that selectively binds to Rsk-3
 CC polypeptide can be used to manufacture reagents for diagnosis of cancer.
 CC Rsk-3 can be used to treat or ameliorate cancer. Rsk-3 inhibitors can
 CC also be used to treat cancer. Vaccines comprising mutant Rsk-3 or nucleic
 CC acids encoding mutant Rsk-3, where the mutant is found in cancer cells,
 CC are useful for treatment of cancer.
 XX SQ Sequence 733 AA;
 Query Match 4.2%; Score 8; DB 20; Length 733;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 128 RRSRSKS 135
 DB 19 RRSRSKS 26
 RESULT 11
 AAW93973
 ID AAW93973 standard; protein; 15 AA.
 AC AAW93973;
 XX 30-JUN-1999 (first entry)
 DT Human IRS-1 and IRS-2 binding inhibitor peptide 1.
 DE Insulin receptor substrate; IRS-1; IRS-2; inhibitor; disease; obesity;
 XX insulin resistance; 14-3-3 protein; inhibitor; screening; diabetes;
 KW diabetic retinopathy; diabetic neuropathy; glucose tolerance;
 KW diabetic nephropathy; hyperinsulinaemia; hyperlipemia; arteriosclerosis;
 KW hypertension; ischaemic heart disease; ischaemic brain disease;
 KW peripheral embolism.
 XX Homo sapiens.
 OS WO9916462-A1.
 XX 08-APR-1999.
 PD 25-SEP-1998; 98WO-JP04293.
 PF 29-SEP-1997; 97JP-0263719.
 XX (DAUC) DAIICHI PHARM CO LTD.
 PA

PI Asano T, Kanda A, Kubo H, Yazaki Y;
 XX WPI: 1999-254929/21.
 DR Treatment of insulin resistance using insulin receptor substrate
 XX binding inhibitor
 PT Disclosure; Page 23; 30pp; Japanese.
 PS This invention describes a method for the treatment of diseases involving
 XX insulin resistance using a substance which inhibits the binding of
 CC insulin receptor substrate to 14-3-3 protein, identified by screening
 CC potential inhibitors for their ability to inhibit this binding. The
 CC composition described in the invention for the treatment of diseases
 CC involving insulin resistance contains as active component, an inhibitor
 CC of the binding of all or part of insulin receptor substrate 1 or 2
 CC (IRS-1, IRS-2) to all or part of 14-3-3 protein. Disorders involving
 CC insulin resistance include diabetes, diabetic retinopathy, diabetic
 CC neuropathy, impaired glucose tolerance, diabetic nephropathy,
 CC hyperinsulinaemia, hyperlipemia, arteriosclerosis, hypertension,
 CC obesity, ischaemic heart disease, ischaemic brain disease and peripheral
 CC embolism.
 XX SQ Sequence 15 AA;
 Query Match 3.6%; Score 7; DB 20; Length 15;
 Best Local Similarity 100.0%; Pred. No. 3.3;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 132 RKSQSS 138
 DB 6 RKSQSS 12
 RESULT 12
 AAG82114
 ID AAG82114 standard; Protein; 129 AA.
 AC AAG82114;
 XX 03-SEP-2001 (first entry)
 DT S. epidermidis open reading frame protein sequence SEQ ID NO:1322.
 DE Staphylococcus epidermidis SRI strain; infection; diagnosis;
 KW vaccination; endocarditis.
 KW Staphylococcus epidermidis.
 XX WO200134809-A2.
 PN 17-MAY-2001.
 PD 09-NOV-2000; 2000WO-US30782.
 PF 09-NOV-1999; 99US-0164258.
 XX (GLAXO) GLAXO GROUP LTD.
 PA Kimmerly WJ;
 XX WPI: 2001-316495/33.
 DR N-PSDB; AAH52964.
 DR Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
 PT useful for vaccinating against infections, e.g. endocarditis -
 XX Claim 18; Page 376-377; 2188pp; English.
 CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
 CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
 CC (I) and (II) can have antibacterial activity and therefore can be used
 CC in vaccination. The nucleic acids (I) may be used to produce the

```

PF 29-OCT-2001; 2001WO-GB04789.
XX
PR 27-OCT-2000; 2000GB-0026333.
PR 24-NOV-2000; 2000GB-0028727.
PR 07-MAR-2001; 2001GB-0005640.
XX
XX (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
XX Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;
PI Tettelin H;
XX
XX WPI; 2002-352536/38.
DR N-PSDB; ABN69287.
XX
XX New Streptococcus protein for the treatment or prevention of infection
PT or disease caused by Streptococcus bacteria, such as meningitis, and
PT for detecting a compound that binds to the protein -
XX
XX Claim 1; Page 3811; 4525pp; English.
XX
XX The invention relates to a protein (ABP25413-ABP30895) from group B
CC Streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins.
XX
XX Sequence 303 AA;
XX
XX Query Match 4.2%; Score 8; DB 23; Length 303;
XX Best Local Similarity 100.0%; Pred. No. 4.4;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
OY 47 TVFGQRES 54
DB 27 TVFGQRES 34
XX
XX RESULT 9
XX AAY30937
XX ID AAY30937 standard; Protein; 733 AA.
XX AC AAY30937;
XX
XX 18-OCT-1999 (first entry)
XX
XX Human rsk3 protein.
XX
XX CD30; antigen; rsk3; ribosomal S6 kinase; pp90rsk3; cell activation;
XX antitumor; antiviral; antibacterial; antiparasitic; anti-allergic;
XX anti-inflammatory; antiproliferative; suppressor; cell death; human;
XX activation-induced proliferation; viral; tumour cell; chemotherapy;
XX cytokine; radiation; treatment; prevention immune dysregulation; blood;
XX allergy; inflammatory disease; transplant rejection; germ cell; mucin;
XX Hodgkin's lymphoma; anaplastic giant cell leukemia; embryonal carcinoma;
XX transplant; bone marrow; cystic fibrosis.
XX
XX Homo sapiens.
XX OS
XX WO940187-A1.
XX PN
XX

```

```

PD 12-AUG-1999.
XX
XX 05-FEB-1999; 99WO-EP00759.
XX
XX 22-DEC-1998; 98DE-1059056.
PR 06-FEB-1998; 98DE-1038967.
XX
XX (ABKE/) ABKEN H.
PA
XX
XX Abken H;
XX
XX WPI; 1999-494292/41.
DR N-PSDB; AAZ09173.
XX
XX New nucleic acids that inhibit expression of the CD30 antigen or
PT ribosomal kinase rsk3, for modulating activity of, or killing, cells
PT expressing these proteins, e.g. tumor cells
XX
XX Claim 10; Page 45-48; 60pp; German.
XX
XX This invention describes novel nucleic acids (I), their derivatives and
CC sequence-specific binding substances that inhibit expression of the CD30
CC antigen or the ribosomal S6 kinase pp90rsk3 (rsk3). The products of the
CC invention have antitumor, antiviral, antibacterial, antiparasitic,
CC anti-allergic, anti-inflammatory and antiproliferative activity. (I)
CC suppress expression of CD30 and/or rsk3 by specific binding to related
CC nucleic acid, so prevent activation of CD30/rsk3-positive cells, restore
CC deregulated production of cellular products (particularly cytokines or
CC other effectors), suppress activation-induced proliferation and viral
CC production, induce cell death and may render tumour cells sensitive to
CC killing by chemotherapeutics, cytokines or radiation. (I) are used to
CC (I) modulate activity of CD30 and/or rsk3-positive cells, and (2) for
CC functional suppression or killing of such cells. Typical applications
CC are: treatment/prevention of immune dysregulation (e.g. in cases of
CC allergy, inflammatory disease, transplant rejection); reduction of viral
CC load (e.g. in infections by human immune deficiency virus, Epstein-Barr
CC virus, hepatitis); elimination of tumours (many of which express CD30,
CC e.g. Hodgkin's lymphoma, anaplastic giant cell leukemia, germ cell
CC tumours and embryonal carcinomas), and treatment of viral, bacterial and
CC parasitic infections. They may also be used to suppress/kill such cells
CC in transplant infections, bone marrow samples, also in blood (for treatment of
CC allergies), and further (I) directed against rsk3 inhibit mucin
CC overexpression induced by Pseudomonas aeruginosa, e.g. in cases of cystic
CC fibrosis. (I) selectively inhibit, or kill, CD30 and/or rsk3-positive
CC cells without significant harm to other cells or tissues. This
CC sequence represents the human rsk3 protein which is used in the method of
CC the invention.
XX
XX Sequence 733 AA;
XX
XX Query Match 4.2%; Score 8; DB 20; Length 733;
XX Best Local Similarity 100.0%; Pred. No. 10;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
OY 128 RRKSRSKS 135
DB 19 RRKSRSKS 26
XX
XX RESULT 10
XX AAY22202
XX ID AAY22202 standard; Protein; 733 AA.
XX
XX AAY22202;
XX
XX 13-SEP-1999 (first entry)
XX
XX p90 ribosomal S6 kinase-3 protein sequence.
XX
XX Rsk-3; p90 ribosomal S6 kinase-3; diagnosis; therapy; ovarian cancer;
XX breast cancer; lymphoma; cancer; vaccine.
XX
XX Homo sapiens.
XX OS

```

ID ABB50067 standard; Protein: 91 AA.
 XX ABB50067;
 AC
 XX
 DT 05-FEB-2002 (first entry)
 XX
 DE Listeria monocytogenes protein #2771.
 XX
 KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
 KW vitamin B12; bacterial infection; disease.
 XX
 OS Listeria monocytogenes.
 XX
 PN WO200177335-A2.
 XX
 PD 18-OCT-2001.
 XX
 PF 11-APR-2001; 2001WO-FR01118.
 XX
 PR 11-APR-2000; 2000FR-0004629.
 XX
 PA (INSP) INST PASTEUR.
 XX
 PI Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
 PI Dussurget O, Chetoui F, Nedjari H, Glaser P, Kunst F, Cossart P;
 PI Daniels J, Goebel W, Kref J, Kuhn M, Ng E, Vazquez-Boland JA;
 PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
 PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
 PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
 PI Madueno E, De Pablos B, Wehlant J, Kaerst U, Entian K, Hauf J;
 PI Rose M, Voss H;
 XX
 WPI: 2002-010914/01.
 DR
 XX
 PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
 PT and prevention of Listeria and related bacterial infections, and
 PT related polypeptides -
 XX
 PS Claim 6; SEQ ID No 2772; 192pp; French.
 XX
 CC The present invention relates to the genome sequence of Listeria
 CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
 CC it are useful for selecting probes and primers for detecting genes in L.
 CC monocytogenes and related organisms, and for studying genetic
 CC polymorphisms and other genomes. The present sequence is a protein
 CC encoded by the genome sequence of the present invention. Proteins
 CC expressed from the genome sequence are useful for raising specific
 CC antibodies, identification of L. monocytogenes and related organisms, and
 CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
 CC B12. The genome sequence and proteins encoded by it are also useful for
 CC selecting compounds that regulate gene expression and cell replication
 CC and modulate L. monocytogenes-related diseases. In addition, the genome
 CC sequence and proteins encoded by it are useful in pharmaceutical and
 CC vaccines compositions for the treatment or prevention of infections by L.
 CC monocytogenes and related organisms.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 91 AA;
 Query Match 5.7%; Score 11; DB 23; Length 91;
 Best Local Similarity 100.0%; Pred. No. 0.0008;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 99 EYCNFGCAVY 109
 Db 76 EYCNFGCAVY 86
 RESULT 7
 ABB53482
 ID ABB53482 standard; Protein: 251 AA.

XX ABB53482;
 AC
 XX
 DT 16-MAY-2002 (first entry)
 XX
 DE Lactococcus lactis protein ybiC.
 XX
 KW Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
 XX
 OS Lactococcus lactis IL1403.
 XX
 PN FR2807446-A1.
 XX
 PD 12-OCT-2001.
 XX
 PF 11-APR-2000; 2000FR-0004630.
 XX
 PR 11-APR-2000; 2000FR-0004630.
 XX
 PA (INRG) INRA INST NAT RECH AGRONOMIQUE.
 XX
 PI Bolotine A, Sorokine A, Renault P, Ehrlich SD;
 XX
 WPI: 2002-043418/06.
 DR
 XX
 PT New nucleotide sequence useful in the identification or Lactococcus
 PT lactis and related species -
 XX
 PS Claim 6; SEQ ID No 184; 2504pp; French.
 XX
 CC The present invention is related to a Lactococcus lactis nucleotide
 CC sequence (ABA90521) and related proteins (ABB53300-ABB55621). The
 CC nucleic acid sequence is useful in the detection and/or amplification of
 CC nucleic acid sequence, particularly to identify Lactococcus lactis or
 CC related species. The proteins of the invention are useful for the
 CC biosynthesis or biodegradation of a composition of interest. The
 CC invention helps research in lactic bacteria, particularly useful in the
 CC production of yogurt and cheese.
 CC Note: The sequence data for this patent is based on equivalent patent
 CC WO200177334 (published 18-OCT-2001) which is available in electronic
 CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 251 AA;
 Query Match 4.2%; Score 8; DB 23; Length 251;
 Best Local Similarity 100.0%; Pred. No. 3.7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 10 LNYNKYPG 17
 Db 10 LNYNKYPG 17
 RESULT 8
 ABB28656
 ID ABB28656 standard; Protein: 303 AA.
 XX
 AC ABB28656;
 XX
 DT 02-JUL-2002 (first entry)
 XX
 DE Streptococcus polypeptide SEQ ID NO 6488.
 XX
 KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 KW group A streptococcus; Streptococcus pyogenes; antibacterial;
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy.
 XX
 OS Streptococcus agalactiae.
 XX
 PN WO200234771-A2.
 XX
 PD 02-MAY-2002.
 XX


```

RESULT 4
AAU37974
ID AAU37974 standard; Protein; 176 AA.
XX AC AAU37974;
XX DT 14-FEB-2002 (first entry)
XX DE Streptococcus pneumoniae cellular proliferation protein #403.
XX KW Antisense; prokaryotic cellular proliferation protein;
XX KW antibiotic; antibacterial; drug design.
XX OS Streptococcus pneumoniae.
XX PN WO200170955-A2.
XX PD 27-SEP-2001.
XX PF 21-MAR-2001; 2001WO-US09180.
XX PR 21-MAR-2000; 2000US-191078P.
XX PR 23-MAY-2000; 2000US-206848P.
XX PR 26-MAY-2000; 2000US-207727P.
XX PR 23-OCT-2000; 2000US-242578P.
XX PR 27-NOV-2000; 2000US-253625P.
XX PR 22-DEC-2000; 2000US-257931P.
XX PR 16-FEB-2001; 2001US-269308P.
XX PA (ELIT-) ELITRA PHARM INC.
XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX PI Yamamoto RT, Xu HH;
XX DR WPI; 2001-611495/70.
XX DR N-PSDB; AAS55833.
XX PT New polynucleotides for the identification and development of
XX PT antibiotics, comprise sequences of antisense nucleic acids -
XX PS Example 3; Seq ID No 13567; 511pp; English.
XX CC The invention relates to antisense inhibitors of genes essential to
XX CC prokaryotic cellular proliferation, their use in identifying the
XX CC genes, their use in the discovery of novel antibiotics, the essential
XX CC genes themselves and the encoded proteins. The prokaryotes used are
XX CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
XX CC invention is also useful for the identification of potential new targets
XX CC for antibiotic development. The antisense nucleic acids can also be used
XX CC to identify proteins used in proliferation, to express these proteins,
XX CC and to obtain antibodies capable of binding to the expressed proteins.
XX CC The proteins can be used to screen compounds in rational drug discovery
XX CC programmes. The antisense nucleic acid sequence is also useful to screen
XX CC for homologous nucleic acids which are required for cell proliferation in
XX CC a wide variety of organisms. The present sequence represents an
XX CC essential prokaryotic cellular proliferation protein.
XX CC Note: The sequence data for this patent did not form part
XX CC of the printed specification, but was obtained in electronic
XX CC format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 176 AA;
Query Match 6.2%; Score 12; DB 22; Length 176;
Best Local Similarity 100.0%; Pred. No. 0.00012;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 69 NEQLRLRGFYKD 80
DB 69 NEQLRLRGFYKD 80
|||||
RESULT 5
ABP25487
ID ABP25487 standard; Protein; 214 AA.
XX AC ABP25487;
XX DT 02-JUL-2002 (first entry)
XX DE Streptococcus polypeptide SEQ ID NO 150.
XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
XX KW group A streptococcus; Streptococcus pyogenes; antibacterial;
XX KW antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX OS Streptococcus pyogenes.
XX PN WO200234771-A2.
XX PD 02-MAY-2002.
XX PF 29-OCT-2001; 2001WO-GB04789.
XX PR 27-OCT-2000; 2000GB-0026333.
XX PR 24-NOV-2000; 2000GB-0028727.
XX PR 07-MAR-2001; 2001GB-0005640.
XX PA (CHIR-) CHIRON SPA.
XX PA (GENO-) INST GENOMIC RES.
XX PI Telford J, Massignani V, Margarit Ros YI, Grandi G, Fraser C;
XX PI Tettelin H;
XX DR WPI; 2002-352536/38.
XX DR N-PSDB; ABN66118.
XX PT New Streptococcus protein for the treatment or prevention of infection
XX PT or disease caused by Streptococcus bacteria, such as meningitis, and
XX PT for detecting a compound that binds to the protein -
XX PS Claim 1; Page 3168; 4525pp; English.
XX CC The invention relates to a protein (ABP25413-ABP30895) from group B
XX CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
XX CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
XX CC the specification. The proteins have antibacterial and antiinflammatory
XX CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
XX CC antibodies that bind (I) are used in the manufacture of medicaments for
XX CC the treatment or prevention of infection or disease caused by
XX CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
XX CC Nucleic acids encoding (I) are used to detect Streptococcus in a
XX CC biological sample. (I) is used to determine whether a compound binds to
XX CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
XX CC used as a vaccine or diagnostic composition. The disease caused by
XX CC Streptococcus that is prevented or treated may be meningitis. Nucleic
XX CC acid encoding (I) may be used to recombinantly produce (I) and may be
XX CC used in gene therapy. Antibodies to (I) are used for affinity
XX CC chromatography, immunoassays, and distinguishing/identifying
XX CC Streptococcus proteins.
XX SQ Sequence 214 AA;
Query Match 6.2%; Score 12; DB 23; Length 214;
Best Local Similarity 100.0%; Pred. No. 0.00015;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 63 IVGDWGNQEURL 74
DB 71 IVGDWGNQEURL 82
|||||
RESULT 6
ABB50067

```


960 2.6 44 22 ABB30911 Peptide #3562 enco
961 2.6 44 22 ABB35277 Peptide #2783 enco
962 2.6 44 22 ABB35386 Peptide #2892 enco
963 2.6 44 22 ABB36094 Peptide #3600 enco
964 2.6 44 22 ABB20720 Protein #2719 enco
965 2.6 44 22 ABB20826 Protein #2823 enco
966 2.6 44 22 ABB21485 Protein #3484 enco
967 2.6 44 22 AAM95781 Human reproductive
968 2.6 44 22 AAM56107 Human brain expres
969 2.6 44 22 AAM56215 Human brain expres
970 2.6 44 22 AAM56876 Human brain expres
971 2.6 44 22 AAM68478 Human bone marrow
972 2.6 44 22 AAM68591 Human bone marrow
973 2.6 44 22 AAM69265 Human bone marrow
974 2.6 44 22 AAM89275 Human immune/haema
975 2.6 44 22 AAC02529 Human polypeptide
976 2.6 44 22 AAC02779 Human polypeptide
977 2.6 44 22 AAM16286 Peptide #2720 enco
978 2.6 44 22 AAM16398 Peptide #2832 enco
979 2.6 44 22 AAM17095 Peptide #3529 enco
980 2.6 44 22 AAM28775 Peptide #2812 enco
981 2.6 44 22 AAM28895 Peptide #2932 enco
982 2.6 44 22 AAM29585 Peptide #3622 enco
983 2.6 44 22 AAM04019 Peptide #2701 enco
984 2.6 44 22 AAM04134 Peptide #2816 enco
985 2.6 44 22 AAM04793 Peptide #3475 enco
986 2.6 44 23 ABG38060 Human peptide enco
987 2.6 44 23 ABG38168 Human peptide enco
988 2.6 44 23 ABG38878 Human peptide enco
989 2.6 45 20 AAY26052 Human Zalphal epit
990 2.6 45 21 AAB16574 Bacteriophage 44AH
991 2.6 45 21 AAG04148 Arabidopsis thalia
992 2.6 45 21 AAG34478 Arabidopsis thalia
993 2.6 45 21 AAY65755 Breast cancer susc
994 2.6 45 22 ABB37435 Peptide #4941 enco
995 2.6 45 22 ABB22734 Protein #4733 enco
996 2.6 45 22 AAU20159 Human novel endocr
997 2.6 45 22 AAM84395 Human immune/haema
998 2.6 45 22 AAB66328 D radiodurans glut
999 2.6 45 23 AAB81765 Human alpha helica
1000 2.6 46 11 AAR05208 Ubiquitin C-termin

ALIGNMENTS

RESULT 1
AAY96805
ID AAY96805 standard; Protein; 192 AA.

XX AC AAY96805;

XX 26-SRP-2000 (first entry)

XX Streptococcus agalactiae pho3-1 protein.

XX Group B Streptococcus; pho3-1; virulence; vaccine; uropathic;
KW gene therapy; veterinary; mastitis; immunostimulant; antibacterial.

XX Streptococcus agalactiae.

XX WO200037646-A2.

XX 29-JUN-2000.

XX 22-DEC-1999; 99WO-GB04377.

XX 22-DEC-1998; 98GB-0028345.

XX 22-DEC-1998; 98GB-0028349.

XX 22-DEC-1998; 98GB-0028350.

XX 22-DEC-1998; 98GB-0028352.

XX 22-DEC-1998; 98GB-0028353.

XX 22-DEC-1998; 98GB-0028354.

PR 22-DEC-1998; 98GB-0028355.
PR 22-DEC-1998; 98GB-0028356.
PR 22-DEC-1998; 98GB-0028357.
PR 22-DEC-1998; 98GB-0028359.
PR 04-JAN-1999; 98GB-0000082.
PR 04-JAN-1999; 98GB-0000083.
PR 04-JAN-1999; 98GB-0000084.
PR 04-JAN-1999; 98GB-0000085.
PR 04-JAN-1999; 98GB-0000086.
PR 28-JAN-1999; 98GB-0001916.
PR 28-JAN-1999; 98GB-0001922.
XX (MICR-) MICROSCIENCE LTD.
PA Hughes MJG, Santangelo JD, Lane JD, Everest P, Feldman R;
XX Moore JC, Wilson RK, Dobson RJ, Dougan G;
XX WPI: 2000-442674/38.
XX N-PSDB; AAA51368.

XX New peptides useful for treatment and prevention of conditions
XX associated with Streptococcal infection are obtained from group B
XX Streptococcus

XX Claim 2: Page 47-48; 63pp; English.

XX The S. agalactiae pho3-1 protein has homology with proteins of unknown
XX function in S. pyogenes, S. pneumoniae, Bacillus subtilis (yutD) and
XX Enterococcus faecalis. The B. subtilis yutD gene is located in a
XX chromosomal region containing genes involved in cell wall synthesis.
XX Peptides derived from group B Streptococcus (Streptococcus agalactiae)
XX proteins encoded by genes pho1-13, pho3-21, pho3-15, pho3-18, pho3-22,
XX pho3-3, pho3-17, pho3-2, pho1-5, pho3-1, pho3-23, pho3-50, pho1-14,
XX pho2-10, pho3-14, pho3-24 and pho3-29. The peptides are useful for
XX screening potential drugs, or for the detection of virulence, and for the
XX manufacture of a medicament for use in the treatment or prevention of
XX infections such as focal infection (including osteomyelitis, septic
XX arthritis, abscesses and endophthalmitis) and urinary tract infections
XX caused by group B Streptococcus. The peptides and vaccines comprising the
XX peptides are useful in treatment of chronic mastitis, especially in cows
XX (i.e. for veterinary purposes).

XX SQ Sequence 192 AA;

XX Query Match 100.0%; Score 192; DB 21; Length 192;

XX Best Local Similarity 100.0%; Pred. No. 8.4e-200;

XX Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRKEVTPPEMLNKNYPGPQFIHFENIVKSDDIIEFQLVINEKSAFDVTVFGQRFSEILKY 60
DB 1 MRKEVTPPEMLNKNYPGPQFIHFENIVKSDDIIEFQLVINEKSAFDVTVFGQRFSEILKY 60

QY 61 DFIVGDWGNQRLRGFYKDASTIRKNSRISRLIEDYIKEYCNFCGAYFVLENNPRDIKF 120
DB 61 DFIVGDWGNQRLRGFYKDASTIRKNSRISRLIEDYIKEYCNFCGAYFVLENNPRDIKF 120

QY 121 DDERPHKRRKRSKSSQSSKTSQTRNNRSQSNANAHFTSKRKDKTKRQERHIKEEQDKEMT 180
DB 121 DDERPHKRRKRSKSSQSSKTSQTRNNRSQSNANAHFTSKRKDKTKRQERHIKEEQDKEMT 180

QY 181 SAKQHLLFVRKN 192
DB 181 SAKQHLLFVRKN 192

RESULT 2
ABP25486
ID ABP25486 standard; Protein; 192 AA.

XX ABP25486;

XX 02-JUL-2002 (first entry)

814	5	2.6	24	23	ABB76466	Mouse IgA inducing	887	5	2.6	36	22	AAE06927	Peptide comprising
815	5	2.6	24	23	ABG42626	Human peptide enco	888	5	2.6	37	20	AAV12015	Human 5' EST secre
816	5	2.6	25	20	AAV27521	E. coli beta'-subu	889	5	2.6	37	22	AAO06370	Human polypeptide
817	5	2.6	25	21	AAE09166	Hepatitis GB virus	890	5	2.6	37	22	AAO07512	Human polypeptide
818	5	2.6	25	23	ABB76462	Bovine IgA inducin	891	5	2.6	37	22	AAO13714	Human polypeptide
819	5	2.6	25	23	ABG62203	Eubacterial DNA Po	892	5	2.6	37	22	AAAB47386	Conjugated peptide
820	5	2.6	25	23	ABG62247	Eubacterial DNA Po	893	5	2.6	37	22	AAAB47387	Conjugated peptide
821	5	2.6	26	15	AAE52512	Human heavy chain	894	5	2.6	37	22	AAAG93072	C glutamic prote
822	5	2.6	26	20	AAV07846	Human secreted pro	895	5	2.6	37	22	AAAB4082	Immunomodulatory p
823	5	2.6	26	22	AAAG5969	Rat SNS2 ion chann	896	5	2.6	37	22	AAE00807	Homoeodomain transc
824	5	2.6	27	23	AAU89448	Insulin/insulin-11	897	5	2.6	38	21	AAE22897	Rat ubiquitin prot
825	5	2.6	28	19	AAV21523	Human ubiquitin B2	898	5	2.6	38	21	AAAB08350	Amino acid sequenc
826	5	2.6	28	20	AAV36437	Fragment of human	899	5	2.6	38	21	AAAY95944	E. coli seryl-tRNA
827	5	2.6	28	21	AAV56397	Mutant ubiquitin-B	900	5	2.6	38	21	AAAG11799	Arabidopsis thalia
828	5	2.6	29	18	AAW10846	MAB anti-HBsAg bin	901	5	2.6	38	21	AAAG23817	Arabidopsis thalia
829	5	2.6	29	19	AAV21524	Human ubiquitin B	902	5	2.6	38	21	AAAG47711	Arabidopsis thalia
830	5	2.6	29	20	AAV27862	Human secreted pro	903	5	2.6	38	22	ABB30388	Peptide #3039 enco
831	5	2.6	29	22	ABE37971	Peptide #5477 enco	904	5	2.6	38	22	ABB35561	Peptide #3067 enco
832	5	2.6	29	22	ABB23215	Protein #5214 enco	905	5	2.6	38	22	ABB39688	Peptide #7194 enco
833	5	2.6	29	22	AAW58602	Human brain expres	906	5	2.6	38	22	ABB20989	Protein #2988 enco
834	5	2.6	29	22	AAW71106	Human bone marrow	907	5	2.6	38	22	AAW56376	Human brain expres
835	5	2.6	29	22	AAW18951	Peptide #5285 enco	908	5	2.6	38	22	AAW60406	Human brain expres
836	5	2.6	29	22	AAW31382	Peptide #5419 enco	909	5	2.6	38	22	AAW68756	Human bone marrow
837	5	2.6	29	22	AAV97753	G. oxydans cytochr	910	5	2.6	38	22	AAW73042	Human bone marrow
838	5	2.6	29	23	ABG40900	Human peptide enco	911	5	2.6	38	22	AAW16579	Peptide #3013 enco
839	5	2.6	30	21	AAV96439	Variant human haem	912	5	2.6	38	22	AAW29062	Peptide #3099 enco
840	5	2.6	30	21	AAV56502	Ubiquitin-B amino	913	5	2.6	38	22	AAW33264	Peptide #7301 enco
841	5	2.6	30	22	ABE37821	Peptide #5327 enco	914	5	2.6	38	22	AAW04392	Peptide #2974 enco
842	5	2.6	30	22	ABB23090	Protein #5089 enco	915	5	2.6	38	22	ABG38338	Human peptide enco
843	5	2.6	30	22	AAW58437	Human brain expres	916	5	2.6	38	23	ABG42886	Human peptide enco
844	5	2.6	30	22	AAW70929	Human bone marrow	917	5	2.6	39	20	AAV11841	Human 5' EST secre
845	5	2.6	30	22	AAW88486	Human immune/haema	918	5	2.6	39	21	AAAG58986	Arabidopsis thalia
846	5	2.6	30	22	AAW18734	Peptide #5168 enco	919	5	2.6	39	21	AAV59607	Arabidopsis thalia
847	5	2.6	30	22	AAW31219	Peptide #5256 enco	920	5	2.6	39	21	AAV66087	BRCA1-associated R
848	5	2.6	30	23	ABG40725	Human peptide enco	921	5	2.6	39	22	ABB40934	Peptide #8440 enco
849	5	2.6	31	19	AAW79678	Synthetic CS198 de	922	5	2.6	39	22	AAW25055	Protein #7054 enco
850	5	2.6	31	22	AAE07321	Human CS 198 pepti	923	5	2.6	39	22	AAW61796	Human brain expres
851	5	2.6	31	22	AAW83189	Human immune/haema	924	5	2.6	39	22	AAW74595	Human bone marrow
852	5	2.6	31	23	ABW44771	Human protective s	925	5	2.6	39	22	AAO07525	Human polypeptide
853	5	2.6	32	16	AAE85036	Peptide rI from th	926	5	2.6	39	22	AAW20355	Peptide #6789 enco
854	5	2.6	32	21	AAW44641	Human secreted pro	927	5	2.6	39	22	AAW34707	Peptide #8744 enco
855	5	2.6	32	21	AAW87774	Interleukin-10 pro	928	5	2.6	39	23	ABG44433	Human peptide enco
856	5	2.6	32	21	AAW87775	Interleukin-10 pro	929	5	2.6	40	13	AAE25790	[Arg16,21,22,25,27
857	5	2.6	32	22	ABW38466	Peptide #5972 enco	930	5	2.6	40	22	ABW41608	Peptide #9114 enco
858	5	2.6	32	22	AAW59077	Human brain expres	931	5	2.6	40	22	ABW25434	Protein #7433 enco
859	5	2.6	32	22	AAW71610	Human bone marrow	932	5	2.6	40	22	AAW62479	Human brain expres
860	5	2.6	32	22	AAW31907	Peptide #5944 enco	933	5	2.6	40	22	AAW75288	Human bone marrow
861	5	2.6	32	23	ABG41422	Human peptide enco	934	5	2.6	40	22	AAW84650	Human immune/haema
862	5	2.6	32	23	ABP26729	Streptococcus poly	935	5	2.6	40	22	AAW20559	Peptide #6993 enco
863	5	2.6	32	23	AAU70404	Human light chain	936	5	2.6	40	22	AAW35401	Peptide #9438 enco
864	5	2.6	33	18	AAW10861	MAB anti-HBsAg bin	937	5	2.6	40	23	ABG44928	Human peptide enco
865	5	2.6	33	22	ABW40890	Peptide #8396 enco	938	5	2.6	41	21	AAW34352	Gene 3 human secre
866	5	2.6	33	22	AAU27374	Novel bone marrow	939	5	2.6	41	22	AAO07432	Human polypeptide
867	5	2.6	33	22	AAW61750	Human brain expres	940	5	2.6	41	22	AAO09925	Human polypeptide
868	5	2.6	33	22	AAW74546	Human bone marrow	941	5	2.6	41	22	AAO13323	Human polypeptide
869	5	2.6	33	22	AAW34658	Peptide #8695 enco	942	5	2.6	41	22	AAW33596	Peptide #7633 enco
870	5	2.6	33	22	AAW47390	Conjugated peptide	943	5	2.6	41	22	AAW77086	Human colon cancer
871	5	2.6	33	22	AAW47391	Conjugated peptide	944	5	2.6	41	23	ABG43250	Human peptide enco
872	5	2.6	34	18	AAW19081	Trypanosoma cruzi	945	5	2.6	42	21	AAW11798	Arabidopsis thalia
873	5	2.6	34	20	AAW96175	IKK-alpha polypept	946	5	2.6	42	21	AAW56941	Arabidopsis thalia
874	5	2.6	34	21	AAW326476	Synthetic peptide	947	5	2.6	42	22	AAW64228	Human brain expres
875	5	2.6	34	21	AAW55417	Arabidopsis thalia	948	5	2.6	43	21	AAW11797	Arabidopsis thalia
876	5	2.6	34	22	AAW47416	Peptide #46 for il	949	5	2.6	43	21	AAW35517	Arabidopsis thalia
877	5	2.6	34	22	AAW92242	Human digestive sy	950	5	2.6	43	22	ABW15364	Human nervous syst
878	5	2.6	34	22	AAO02683	Human polypeptide	951	5	2.6	43	22	AAW86434	Human immune/haema
879	5	2.6	35	18	AAW15127	Phytase N-terminal	952	5	2.6	43	22	AAO07515	Human polypeptide
880	5	2.6	35	21	AAW57169	Human prostate can	953	5	2.6	43	22	AAO07612	Human polypeptide
881	5	2.6	35	21	AAW59361	Arabidopsis thalia	954	5	2.6	43	23	ABW65617	Human breast speci
882	5	2.6	35	22	AAW47388	Conjugated peptide	955	5	2.6	44	15	AAW54012	Characteristic pro
883	5	2.6	35	22	AAW47389	Conjugated peptide	956	5	2.6	44	21	AAW52537	Helicobacter pylor
884	5	2.6	35	22	AAW47392	Conjugated peptide	957	5	2.6	44	22	ABW96312	Human testicular a
885	5	2.6	36	21	AAW19327	Amino acid sequenc	958	5	2.6	44	22	ABG23704	Novel human diagn
886	5	2.6	36	22	ABB03530	Human musculoskele	959	5	2.6	44	22	ABB30222	Peptide #2873 enco

668	5	2.6	9	23	ABB94488	CTL epitope HLA pe	741	15	23	AAU86188	Oestradiol activat
669	5	2.6	9	23	ABB94917	CTL epitope HLA pe	742	15	23	ABO44468	Human transcriptio
670	5	2.6	10	16	AAW21559	Corticotropin rele	743	16	19	AAW84025	Human CYP3A4 speci
671	5	2.6	10	18	AAW15482	Human p100 protein	744	16	20	AAW97301	Peptide which indu
672	5	2.6	10	19	AAW44362	Breast cancer-asso	745	16	21	AAW65612	Oestrogen receptor
673	5	2.6	10	21	AAW27127	Human CASB619 prot	746	16	22	AAU08975	Human CDC25A inhib
674	5	2.6	10	22	AAW94052	Human complementar	747	16	23	ABB88202	C generalis mu-con
675	5	2.6	10	22	AAW94054	Human complementar	748	16	23	ABB88204	C generalis mu-con
676	5	2.6	10	22	AAW96450	Human complementar	749	16	23	ABB88427	C generalis mu-con
677	5	2.6	10	22	AAW844350	Arabidopsis thalia	750	16	23	ABB88428	C generalis mu-con
678	5	2.6	10	22	AAW86157	Saccharomyces cere	751	16	23	ABB88429	C generalis mu-con
679	5	2.6	10	22	AAW88110	Saccharomyces cere	752	16	23	AAU86349	Oestrogen receptor
680	5	2.6	10	22	AAW88111	Saccharomyces cere	753	16	23	AAU86349	DNA repair protein
681	5	2.6	10	22	AAU03225	Fruit fly G protei	754	16	23	AAU86349	Peptide that mimic
682	5	2.6	10	22	AAU03235	Fruit fly G protei	755	16	23	AAU86349	Streptococcus pneu
683	5	2.6	10	22	AAU03872	G protein-coupled	756	16	23	AAU86349	Mutant ubiquitin-B
684	5	2.6	10	22	AAW98714	Human breast cance	757	16	23	AAU86349	Human FAS partial
685	5	2.6	10	22	AAW30665	Neuropeptide F (NP	758	16	23	AAU86349	Ig kappa chain pos
686	5	2.6	10	22	AAW50005	Human D40 associat	759	16	23	AAU86349	Human CYP3A4 speci
687	5	2.6	10	23	ABB94564	CTL epitope HLA pe	760	16	23	AAU86349	D. melanogaster SL
688	5	2.6	10	23	ABB94728	CTL epitope HLA pe	761	16	23	AAU86349	D. melanogaster SL
689	5	2.6	10	23	ABB94856	CTL epitope HLA pe	762	16	23	AAU86349	Human CYP3A4 speci
690	5	2.6	10	23	ABB94962	CTL epitope HLA pe	763	16	23	AAU86349	Hepatitis GB virus
691	5	2.6	11	13	AAW28088	Cell-to-cell bindi	764	16	23	AAU86349	Alpha D peptide de
692	5	2.6	11	18	AAW11809	Fas ligand antigen	765	16	23	AAU86349	Conjugated peptide
693	5	2.6	11	19	AAW84023	Human CYP3A4 speci	766	16	23	AAU86349	Peptide derived fr
694	5	2.6	11	19	AAW84028	Human CYP3A4 speci	767	16	23	AAU86349	Mouse myosin heavy
695	5	2.6	11	22	AAW65972	Rat SNS1 ion chann	768	16	23	AAU86349	C-terminus of ubiq
696	5	2.6	12	13	AAW20971	Beta-2 integrin pe	769	16	23	AAU86349	Antigenic peptide
697	5	2.6	12	15	AAW67109	Antifertility pept	770	16	23	AAU86349	Cytotoxic T lympho
698	5	2.6	12	17	AAW02057	Human beta2 integr	771	16	23	AAU86349	T-cell stimulatory
699	5	2.6	12	19	AAW84027	Human CYP3A4 speci	772	16	23	AAU86349	Alpha D peptide de
700	5	2.6	12	21	AAW98441	Alpha D peptide de	773	16	23	AAU86349	Peptide #10416 enc
701	5	2.6	12	23	AAU10818	Human Type I GnRH-	774	16	23	AAU86349	Human albumin-bind
702	5	2.6	12	23	ABB74483	DNA repair protein	775	16	23	AAU86349	Protein #8163 enco
703	5	2.6	13	17	AAW02560	Calcineurin-bindin	776	16	23	AAU86349	Human brain expres
704	5	2.6	13	21	AAW27829	Human secreted pro	777	16	23	AAU86349	Peptide #10767 enc
705	5	2.6	13	21	AAW14896	Human calcineurin-	778	16	23	AAU86349	Human bone marrow
706	5	2.6	13	21	AAW98442	Alpha D peptide de	779	16	23	AAU86349	Peptide #10767 enc
707	5	2.6	14	20	AAW00930	Prostate-tumour de	780	16	23	AAU86349	Human peptide enco
708	5	2.6	14	21	AAW66097	BRCA1-associated R	781	16	23	AAU86349	T-cell receptor pe
709	5	2.6	14	22	AAW65973	Rat SNS1 ion chann	782	16	23	AAU86349	T-cell receptor pe
710	5	2.6	14	22	AAW68666	Human peptide #141	783	16	23	AAU86349	T-cell receptor pe
711	5	2.6	14	22	AAW68640	Human peptide #215	784	16	23	AAU86349	Human CYP3A4 speci
712	5	2.6	14	22	AAW97744	Human peptide #708	785	16	23	AAU86349	Peptide #5108 enco
713	5	2.6	14	22	AAW63136	Peptide from B4-5	786	16	23	AAU86349	zinc finger helica
714	5	2.6	14	22	AAW63136	Human osteopontin	787	16	23	AAU86349	Human osteopontin
715	5	2.6	14	23	AAW76045	Ig kappa chain pos	788	16	23	AAU86349	Insulin/insulin-li
716	5	2.6	15	15	AAW47005	Chloroperoxidase t	789	16	23	AAU86349	T cell receptor im
717	5	2.6	15	16	AAW81874	Biologically activ	790	16	23	AAU86349	T cell receptor im
718	5	2.6	15	16	AAW72034	Calcineurin bindin	791	16	23	AAU86349	T cell receptor im
719	5	2.6	15	17	AAW02562	Calcineurin bindin	792	16	23	AAU86349	T cell receptor im
720	5	2.6	15	17	AAW02561	Calcineurin bindin	793	16	23	AAU86349	T cell receptor im
721	5	2.6	15	18	AAW11758	T-cell receptor pe	794	16	23	AAU86349	T cell receptor im
722	5	2.6	15	18	AAW11759	T-cell receptor pe	795	16	23	AAU86349	T cell receptor im
723	5	2.6	15	19	AAW20433	Human microtubule	796	16	23	AAU86349	Human CYP3A4 speci
724	5	2.6	15	19	AAW84021	Human CYP3A4 speci	797	16	23	AAU86349	E. coli beta-subun
725	5	2.6	15	21	AAW13482	C. pneumoniae Cp-S	798	16	23	AAU86349	Hepatitis B virus
726	5	2.6	15	21	AAW14897	Calcineurin-bindin	799	16	23	AAU86349	Novel human diagno
727	5	2.6	15	21	AAW14898	Calcineurin-bindin	800	16	23	AAU86349	Human cytochrome p
728	5	2.6	15	21	AAW13726	C-terminal peptide	801	16	23	AAU86349	A. thaliana SRP30
729	5	2.6	15	21	AAW66098	BRCA1-associated R	802	16	23	AAU86349	Human protective s
730	5	2.6	15	21	AAW65464	Estradiol activate	803	16	23	AAU86349	Peptide #2 used to
731	5	2.6	15	22	AAW99415	Vaccine related MH	804	16	23	AAU86349	E. coli beta-subun
732	5	2.6	15	22	AAW83158	Chlamydia Cp-Swib	805	16	23	AAU86349	Human secreted pro
733	5	2.6	15	22	AAW64800	Human zinc finger	806	16	23	AAU86349	Peptide #6998 enco
734	5	2.6	15	23	AAW67791	Human ADPI tryptic	807	16	23	AAU86349	Human brain expres
735	5	2.6	15	23	AAW67986	Human ADPI tryptic	808	16	23	AAU86349	Human bone marrow
736	5	2.6	15	23	AAW81761	Human alpha helica	809	16	23	AAU86349	ERA binding domain
737	5	2.6	15	23	AAW62301	Eubacterial DNA po	810	16	23	AAU86349	Human interleukin-
738	5	2.6	15	23	AAW81216	Human glutaryl-CoA	811	16	23	AAU86349	Human interleukin-
739	5	2.6	15	23	AAW08849	Human non-insulin	812	16	23	AAU86349	Fluorescently labe
740	5	2.6	15	23	AAW94129	Chlamydia peptide	813	16	23	AAU86349	Human IgA inducing

522	6	3.1	737	22	ABG10283	Novel human diagno	595	6	3.1	1020	22	ABBS58971	Drosophila melanog
523	6	3.1	747	23	ABB90996	Herbicidally activ	596	6	3.1	1029	14	AAR38862	GC-A. Rattus ratt
524	6	3.1	752	20	AAW97799	Streptococcus pneu	597	6	3.1	1033	23	ABBS4120	Drosophila melanog
525	6	3.1	752	20	AAW74407	C. trachomatis gld	598	6	3.1	1033	23	AAU82717	Amino acid sequenc
526	6	3.1	752	22	AAAB4212	Amino acid sequenc	599	6	3.1	1035	23	ABB66062	Drosophila melanog
527	6	3.1	752	22	AAAB8213	Glycogen phosphory	600	6	3.1	1036	23	ABB46886	Listeria monocytog
528	6	3.1	752	23	AAAG79339	Human SR-cyp prote	601	6	3.1	1060	22	AAE09313	Sso/Tag DNA polyme
529	6	3.1	754	23	AAE24596	C glutamicum prote	602	6	3.1	1061	23	AAU11280	Human atrionatriur
530	6	3.1	755	22	AAAG90181	Human polyypeptide	603	6	3.1	1061	23	AAU11281	Human atrionatriur
531	6	3.1	757	22	AAAG39928	Drosophila melanog	604	6	3.1	1071	21	AAG36303	Arabidopsis thalia
532	6	3.1	759	22	ABB68975	Human LP protein L	605	6	3.1	1071	23	ABP27560	Streptococcus poly
533	6	3.1	759	23	AAAG7037	Arabidopsis thalia	606	6	3.1	1072	21	AAG36302	Arabidopsis thalia
534	6	3.1	761	21	AAAG20796	Human Che-2 protei	607	6	3.1	1075	20	AAV35498	C. pneumoniae prot
535	6	3.1	764	23	AAU81523	Arabidopsis thalia	608	6	3.1	1075	22	ABB11783	Human ANP-A recept
536	6	3.1	765	21	AAAG54100	Herbicidally activ	609	6	3.1	1098	22	ABG20355	Novel human diagno
537	6	3.1	773	23	ABB91850	Novel human secret	610	6	3.1	1119	22	ABBS97825	Thermus aquaticus
538	6	3.1	781	22	AAU28171	Soybean lysine ket	611	6	3.1	1119	23	AAE14716	Thermus aquaticus
539	6	3.1	784	20	AAW87760	Drosophila melanog	612	6	3.1	1157	22	ABB63351	Drosophila melanog
540	6	3.1	787	22	ABB58149	Novel human diagno	613	6	3.1	1187	16	AAAG66451	AF-4 protein (enco
541	6	3.1	787	22	ABG13669	Human phosphatidyl	614	6	3.1	1203	22	AAW79264	Human protein SEQ
542	6	3.1	787	23	AAE14673	Human PRO protein,	615	6	3.1	1210	16	AAAG66450	AF-4 protein (enco
543	6	3.1	787	23	AAE93636	Rat Ras signalling	616	6	3.1	1217	22	AAW52358	Putative TBP inter
544	6	3.1	795	21	AAAY70962	Human Ras signalli	617	6	3.1	1305	22	ABBS67631	Drosophila melanog
545	6	3.1	797	21	AAAY70963	Drosophila melanog	618	6	3.1	1322	22	ABBS60275	Drosophila melanog
546	6	3.1	805	22	ABBS59621	Arabidopsis thalia	619	6	3.1	1327	22	ABBS58539	Drosophila melanog
547	6	3.1	810	21	AAAG20795	Human cell cycle a	620	6	3.1	1338	17	AAW91247	VIP2A(a)-VIP1A(a)
548	6	3.1	812	22	AAAG60495	Arabidopsis thalia	621	6	3.1	1338	18	AAW19520	Maize optimised-B
549	6	3.1	815	21	AAAG54099	Arabidopsis thalia	622	6	3.1	1338	19	AAW46731	VIP2A(a)/VIP1A(a)
550	6	3.1	824	22	AAAG98907	E. coli growth and	623	6	3.1	1343	18	AAW31866	Mouse metastasis-a
551	6	3.1	826	21	AAAY95050	Candida albicans p	624	6	3.1	1346	17	AAW91245	VIP2A(a) and VIP1A
552	6	3.1	826	23	ABB91392	Herbicidally activ	625	6	3.1	1346	18	AAW19513	B. cereus VIP1A(a)
553	6	3.1	827	22	AAU35764	Helicobacter pylor	626	6	3.1	1346	19	AAW46723	VIP1A(a)/VIP2A(a)
554	6	3.1	829	17	AAAG86578	Autotaxin derived	627	6	3.1	1363	22	ABBS64266	Drosophila melanog
555	6	3.1	831	19	AAW54372	Carboxydothermus h	628	6	3.1	1443	22	ABG25528	Novel human diagno
556	6	3.1	844	22	ABB63132	Drosophila melanog	629	6	3.1	1464	22	ABBS05043	Novel human diagno
557	6	3.1	850	23	AAW47574	Herbicidally cell cy	630	6	3.1	1477	16	AAW67691	S. cerevisiae scav
558	6	3.1	852	23	AAW92825	Mutant thermotabl	631	6	3.1	1477	18	AAW10424	Saccharomyces cere
559	6	3.1	856	13	AAAG23170	Mutant thermotabl	632	6	3.1	1543	22	AAW06819	Fumonisin-resistan
560	6	3.1	868	20	AAAY00935	A. degensii DNA po	633	6	3.1	1547	22	ABBS69373	Drosophila melanog
561	6	3.1	868	23	AAE22112	Ammonifex degensii	634	6	3.1	1572	23	ABBS97562	Drosophila melanog
562	6	3.1	869	22	ABBS2777	Escherichia coli p	635	6	3.1	1577	23	ABBS99387	Novel human protei
563	6	3.1	882	15	AAAG3996	Thermophilic DNA p	636	6	3.1	1605	22	ABBS70375	Drosophila melanog
564	6	3.1	882	15	AAAG5749	Alpha-DNA polymera	637	6	3.1	1850	22	ABBS72210	Drosophila melanog
565	6	3.1	892	13	AAAG23169	Mutant thermotabl	638	6	3.1	1898	20	AAV30795	Modified chicken v
566	6	3.1	892	13	AAAG23169	Taf DNA polymerase	639	6	3.1	1920	22	ABBS64441	A human trichohyal
567	6	3.1	895	21	AAAG24233	Arabidopsis thalia	640	6	3.1	2042	23	AAE24139	Drosophila melanog
568	6	3.1	903	20	AAAY28713	Detargeted yeast m	641	6	3.1	2052	21	AAAB08634	Human kinase (PKIN
569	6	3.1	904	22	ABB68734	Drosophila melanog	642	6	3.1	2183	22	AAG84988	A murine phosphati
570	6	3.1	906	20	AAAY28702	Yeast multifunctio	643	6	3.1	2183	22	AAG84988	Shrimp white spot
571	6	3.1	906	20	AAAY28712	Mutant yeast multi	644	6	3.1	2387	22	AAU01183	Rat glutamate tran
572	6	3.1	906	22	AAAB20180	Candida tropicalis	645	6	3.1	2390	20	AAV05494	BetaIII spectrin p
573	6	3.1	906	22	AAAB20181	C tropicalis multi	646	6	3.1	2400	22	ABG20278	Novel human diagno
574	6	3.1	906	22	AAAB20182	C. tropicalis mult	647	6	3.1	2415	22	ABG20279	Novel human diagno
575	6	3.1	906	22	AAAB20183	C. tropicalis mult	648	6	3.1	2424	22	ABBS58924	Drosophila melanog
576	6	3.1	915	17	AAAB65996	A2038 autotaxin pr	649	6	3.1	2451	22	ABBS71574	Drosophila melanog
577	6	3.1	915	22	AAAY71987	Human melanoma au	650	6	3.1	2618	22	ABG02135	Novel human diagno
578	6	3.1	921	22	AAAG91905	C glutamicum prote	651	6	3.1	2622	22	ABG06418	Novel human diagno
579	6	3.1	925	23	ABB91769	Herbicidally activ	652	6	3.1	2681	22	ABG29184	Novel human diagno
580	6	3.1	931	22	ABG23862	Novel human diagno	653	6	3.1	2803	23	ABBS08161	Human cytoskeleton
581	6	3.1	938	22	ABBS0562	Drosophila melanog	654	6	3.1	2958	22	ABBS71368	Drosophila melanog
582	6	3.1	950	22	AAU36368	Pseudomonas aerugi	655	6	3.1	3256	21	AAV50976	Human cell cycle p
583	6	3.1	956	21	AAAG42432	Arabidopsis thalia	656	6	3.1	3309	22	ABBS64158	Drosophila melanog
584	6	3.1	959	22	ABBS6368	Drosophila melanog	657	6	3.1	5533	22	ABBS65772	Drosophila melanog
585	6	3.1	960	22	ABBS58434	Drosophila melanog	658	6	3.1	5560	22	ABBS71160	Drosophila melanog
586	6	3.1	970	22	ABBS58624	Drosophila melanog	659	6	3.1	7107	22	ABBS58144	Drosophila melanog
587	6	3.1	982	22	ABBS62346	Drosophila melanog	660	5	2.6	5	20	AAW98057	Streptococcus agal
588	6	3.1	983	22	AAW74684	Human protease and	661	5	2.6	5	23	AAW51068	S. agalactiae sorta
589	6	3.1	985	21	AAG36304	Arabidopsis thalia	662	5	2.6	5	23	AAW52885	Plasmodium falci
590	6	3.1	987	22	AAAB4633	Amino acid sequenc	663	5	2.6	7	22	ABBS1561	Zif268 zinc finger
591	6	3.1	1002	21	AAG42431	Arabidopsis thalia	664	5	2.6	7	23	ABBS74582	Transcription fact
592	6	3.1	1013	21	AAE11053	Amino acid sequenc	665	5	2.6	7	23	ABBS74701	Transcription fact
593	6	3.1	1013	22	AAE11053	Ashbya gossypii Gr	666	5	2.6	8	23	ABBS74583	Transcription fact
594	6	3.1	1014	23	ABB08001	Human lipid metabo	667	5	2.6	9	21	ABBS26189	Human CASB619 prot

376	6	3.1	438	21	AAB03421	Soybean putative c	449	6	3.1	556	22	ABB59543	Drosophila melanog
377	6	3.1	439	21	AAV53797	Amino acid sequenc	450	6	3.1	558	21	AAB23175	Chicory germacrene
378	6	3.1	444	21	AAV19098	Protein encoded by	451	6	3.1	558	22	ABB64967	Drosophila melanog
379	6	3.1	444	23	ABP27271	Streptococcus poly	452	6	3.1	558	23	ABP27849	Streptococcus poly
380	6	3.1	444	23	ABB91258	Herbicidally activ	453	6	3.1	560	22	AAU22961	Novel human enzyme
381	6	3.1	444	23	ABB91259	Herbicidally activ	454	6	3.1	568	22	ABB64744	Drosophila melanog
382	6	3.1	446	19	AAW60232	Bacillus thuringie	455	6	3.1	568	22	AAE10823	Human gene 2 encod
383	6	3.1	446	21	AAV59279	WAR toxin from B.	456	6	3.1	571	22	AAV89879	Thrombomucin, Uni
384	6	3.1	449	17	AAV91249	Vip2A(a) protein w	457	6	3.1	571	23	AAO17248	A thaliana RKSII su
385	6	3.1	449	18	AAV19519	Maize optimised-B.	458	6	3.1	576	22	ABB62520	Drosophila melanog
386	6	3.1	449	19	AAW46730	Vip2A(a) with a va	459	6	3.1	578	22	ABB59076	Drosophila melanog
387	6	3.1	452	21	AAG32061	Arabidopsis thalia	460	6	3.1	579	23	AAO17257	A thaliana RKSIII
388	6	3.1	454	21	AAV41897	Human OREF ORF1661	461	6	3.1	579	23	AAO17260	A thaliana recepto
389	6	3.1	456	21	AAG51360	Arabidopsis thalia	462	6	3.1	580	23	AAO17311	A thaliana recepto
390	6	3.1	458	18	AAW14592	Streptococcus pneu	463	6	3.1	580	22	AAU30814	Novel human secret
391	6	3.1	462	17	AAV91238	B. cereus Vip2A(a)	464	6	3.1	588	23	AAO17251	A thaliana RKSII s
392	6	3.1	462	18	AAV19508	B. cereus Vip2A(a)	465	6	3.1	591	23	AAO17256	A thaliana RKSIII
393	6	3.1	462	19	AAW46710	Vegetative insecti	466	6	3.1	592	23	AAO17258	A thaliana RKSIII
394	6	3.1	463	22	ABB64109	Drosophila melanog	467	6	3.1	596	23	AAO17271	A thaliana recepto
395	6	3.1	466	22	AAV95125	Human protein sequ	468	6	3.1	597	23	AAO17259	A thaliana recepto
396	6	3.1	466	23	AAU76040	Human ADAM-H9 prot	469	6	3.1	598	21	AAG30868	Arabidopsis thalia
397	6	3.1	467	21	AAG30869	Arabidopsis thalia	470	6	3.1	600	11	AAR03922	M.hyponeumonlae HS
398	6	3.1	467	22	ABB50210	Human transcriptio	471	6	3.1	600	14	AAV43003	Mycoplasma 74.5kd
399	6	3.1	472	20	AAV41722	Human PRO871 prote	472	6	3.1	600	23	AAU82971	S. cerevisiae TPB1
400	6	3.1	472	21	AAV44278	Human PRO871 (UNQ4	473	6	3.1	603	16	AAV74620	Human lung tumour
401	6	3.1	472	22	AAU23072	Human PRO polypept	474	6	3.1	607	20	AAV14149	DeltaChy DNA polym
402	6	3.1	472	22	AAV66865	Human peptidyl-pro	475	6	3.1	607	20	AAV14300	DeltaChy DNA polym
403	6	3.1	472	23	ABB95466	Human angioogenesis	476	6	3.1	607	23	AAO17316	A thaliana recepto
404	6	3.1	472	23	ABB84860	Human PRO871 prote	477	6	3.1	610	21	AAG36219	Arabidopsis thalia
405	6	3.1	473	21	AAV59829	Soybean histone de	478	6	3.1	612	21	AAV51362	Arabidopsis thalia
406	6	3.1	473	14	AAV41824	Methylobacillus g1	479	6	3.1	615	21	AAG36218	Arabidopsis thalia
407	6	3.1	477	22	AAU69456	Human purified sec	480	6	3.1	615	21	AAG51359	Arabidopsis thalia
408	6	3.1	482	23	ABP39464	Staphylococcus epi	481	6	3.1	617	23	ABB93952	Herbicidally activ
409	6	3.1	485	21	AAG49236	Arabidopsis thalia	482	6	3.1	619	22	ABB66630	Drosophila melanog
410	6	3.1	485	22	AAE09718	Human ubiquitin ca	483	6	3.1	621	14	AAV41230	Taurine transporte
411	6	3.1	494	19	AAV7017	Arabidopsis thalia	484	6	3.1	621	22	ABB64486	Drosophila melanog
412	6	3.1	495	21	AAG51363	Arabidopsis thalia	485	6	3.1	621	22	ABB64485	Amino acid sequenc
413	6	3.1	501	22	AAU33796	Staphylococcus aur	486	6	3.1	622	22	AAV94338	Human protein sequ
414	6	3.1	504	22	AAU36910	Staphylococcus aur	487	6	3.1	622	22	AAV95098	Human protein sequ
415	6	3.1	507	16	AAW01499	70 kD cellobiohydr	488	6	3.1	622	23	AAU74330	Human cytoskeleton
416	6	3.1	508	22	AAE09309	Sulfobolus sulfata	489	6	3.1	624	23	AAE15257	Human RNA metaboli
417	6	3.1	512	22	ABB65259	Drosophila melanog	490	6	3.1	625	19	AAW47023	Arabidopsis thalia
418	6	3.1	512	22	AAU02059	Synthetic human mu	491	6	3.1	625	23	AAO17246	A thaliana recepto
419	6	3.1	520	21	AAV19740	Arabidopsis acyl C	492	6	3.1	625	23	AAO17254	A thaliana RKSIII
420	6	3.1	520	21	AAV94512	Arabidopsis diacyl	493	6	3.1	625	23	ABB91554	Herbicidally activ
421	6	3.1	520	21	AAV94524	A. thaliana diacyl	494	6	3.1	628	21	AAG30867	Arabidopsis thalia
422	6	3.1	520	21	AAG32060	Arabidopsis thalia	495	6	3.1	628	22	AAV41714	Human polypeptide
423	6	3.1	520	21	AAV35807	Arabidopsis thalia	496	6	3.1	628	23	AAO17255	A thaliana RKSIII
424	6	3.1	520	21	AAV96853	A. thaliana diacyl	497	6	3.1	628	23	AAO17269	A thaliana recepto
425	6	3.1	520	21	AAV54143	Acyl-CoA:cholester	498	6	3.1	628	23	ABB91241	Herbicidally activ
426	6	3.1	520	22	AAU00462	Arabidopsis thalia	499	6	3.1	630	21	AAV25103	Pinus radiata cell
427	6	3.1	520	23	ABB91743	Herbicidally activ	500	6	3.1	632	23	ABB91401	Herbicidally activ
428	6	3.1	522	22	ABB63031	Drosophila melanog	501	6	3.1	632	23	ABB07638	Human speckled l10
429	6	3.1	522	22	ABB65580	Drosophila melanog	502	6	3.1	640	20	ABB61283	Drosophila melanog
430	6	3.1	523	21	AAV31084	Arabidopsis thalia	503	6	3.1	650	20	AAV14962	Amino acid sequenc
431	6	3.1	523	23	ABB93214	Herbicidally activ	504	6	3.1	650	22	AAV83262	C elegans FAIPA SE
432	6	3.1	524	21	AAG32057	Arabidopsis thalia	505	6	3.1	650	22	AAV83274	C elegans FAIPA SE
433	6	3.1	524	21	AAV49211	Arabidopsis thalia	506	6	3.1	660	22	ABB67711	Drosophila melanog
434	6	3.1	524	23	ABB91744	Herbicidally activ	507	6	3.1	661	22	ABB60769	Drosophila melanog
435	6	3.1	527	23	ABB92687	Herbicidally activ	508	6	3.1	665	22	ABB64312	Drosophila melanog
436	6	3.1	528	22	AAV78918	Human protein SEQ	509	6	3.1	679	22	ABG18804	Novel human diagno
437	6	3.1	529	23	ABP43478	Human secreted pro	510	6	3.1	680	22	ABB68821	Drosophila melanog
438	6	3.1	530	21	AAG36220	Arabidopsis thalia	511	6	3.1	689	23	ABB07636	Human speckled l10
439	6	3.1	540	22	ABB11354	Human Ser/Arg rich	512	6	3.1	690	22	ABB63142	Drosophila melanog
440	6	3.1	542	22	AAV94243	Human protein sequ	513	6	3.1	702	22	AAV46724	S. solfatarius DN
441	6	3.1	542	22	AAV95042	Human protein sequ	514	6	3.1	709	22	ABB56396	TSHR-GS-alpha fusi
442	6	3.1	542	23	ABB08186	Human nucleotide r	515	6	3.1	718	22	ABB58626	Drosophila melanog
443	6	3.1	542	23	ABB97456	Novel human protei	516	6	3.1	720	22	ABB58376	Drosophila melanog
444	6	3.1	542	23	AAU81996	Human secreted pro	517	6	3.1	722	12	AAV10055	Cephalosporin C am
445	6	3.1	547	21	AAV96854	A. thaliana AS11 m	518	6	3.1	725	22	AAV99036	Human somatostatin
446	6	3.1	549	22	ABB50168	Human transcriptio	519	6	3.1	727	20	AAV43523	Human CCCYC-bindin
447	6	3.1	552	23	ABB91328	Herbicidally activ	520	6	3.1	728	20	AAV43520	Chicken CCTC-bind
448	6	3.1	553	19	AAW47013	Daucus carota SERK	521	6	3.1	736	23	ABB57361	Mouse ischaemic co

230	6	3.1	283	22	AAB95629	Human protein sequ	303	6	3.1	355	23	ABB09277	G protein-coupled
231	6	3.1	283	23	ABG61546	Human transporter	304	6	3.1	355	23	AAU79335	Human inhibitory G
232	6	3.1	284	22	AAB76732	Corynebacterium gl	305	6	3.1	359	22	ABB67482	Drosophila melanog
233	6	3.1	284	22	AAB76734	Corynebacterium gl	306	6	3.1	359	22	ABB52703	Escherichia coli p
234	6	3.1	286	20	AAY32172	Chlamydia trachoma	307	6	3.1	361	22	AAI22206	Biohytm marker p
235	6	3.1	287	21	AAB36285	Human protein frag	308	6	3.1	361	23	ABB54428	Lactococcus lactis
236	6	3.1	287	22	AAM39480	Human polypeptide	309	6	3.1	368	21	AAG32059	Arabidopsis thalia
237	6	3.1	290	21	AAG48998	Arabidopsis thalia	310	6	3.1	369	17	AAW06822	Turkey herpes viru
238	6	3.1	291	23	AAE15249	Human RNA metaboli	311	6	3.1	369	21	AAC34421	Arabidopsis thalia
239	6	3.1	294	11	AAE06435	Halacetate dehalo	312	6	3.1	369	21	AAC37344	Arabidopsis thalia
240	6	3.1	294	21	AG05510	Arabidopsis thalia	313	6	3.1	372	20	AAI07102	Colon cancer assoc
241	6	3.1	299	21	AAG49929	Arabidopsis thalia	314	6	3.1	375	22	AAW40033	Human polypeptide
242	6	3.1	299	22	AAG56613	Amino acid sequenc	315	6	3.1	378	22	AAAB60643	Arabidopsis thalia
243	6	3.1	301	21	AAB50380	Human uncoupling p	316	6	3.1	380	22	AAB94780	Human protein sequ
244	6	3.1	301	23	AAE16767	Human transporter	317	6	3.1	383	19	AAW98461	H. pylori GHPO 718
245	6	3.1	303	21	AG09218	Arabidopsis thalia	318	6	3.1	385	20	AAW78473	Human 68772 protei
246	6	3.1	305	20	AAI37455	Protein involved i	319	6	3.1	385	21	AAAB41850	Human ORFX ORF1614
247	6	3.1	306	21	AAE54149	Human pancreatic c	320	6	3.1	387	22	ABB59239	Drosophila melanog
248	6	3.1	308	23	ABP30403	Streptococcus poly	321	6	3.1	391	21	AAE37343	Arabidopsis thalia
249	6	3.1	310	22	ABG04751	Novel human diagno	322	6	3.1	391	22	AAE12467	Tobacco translatio
250	6	3.1	312	21	AG09217	Arabidopsis thalia	323	6	3.1	392	22	ABE58291	Drosophila melanog
251	6	3.1	312	21	AG11045	Arabidopsis thalia	324	6	3.1	392	22	ABG09960	Novel human diagno
252	6	3.1	312	21	AG40400	Arabidopsis thalia	325	6	3.1	393	21	AG51361	Arabidopsis thalia
253	6	3.1	312	22	AG71710	Human olfactory re	326	6	3.1	394	22	ABG12146	Novel human diagno
254	6	3.1	312	22	AG72812	Human olfactory re	327	6	3.1	395	21	AG49780	Arabidopsis thalia
255	6	3.1	312	22	AG72978	Olfactory receptor	328	6	3.1	395	23	ABW93740	Herbicidally activ
256	6	3.1	312	23	ABW09888	Herbicidally activ	329	6	3.1	396	19	AAW80682	S. pneumoniae prot
257	6	3.1	314	22	ABE61195	Human testicular a	330	6	3.1	397	23	ABE92240	Herbicidally activ
258	6	3.1	314	22	AAW95512	Human reproductive	331	6	3.1	399	22	ABE63309	Drosophila melanog
259	6	3.1	314	23	ABW47431	Listeria monocytog	332	6	3.1	399	22	AAU17440	Novel signal trans
260	6	3.1	315	22	AG72979	Olfactory receptor	333	6	3.1	401	23	ABW92241	Herbicidally activ
261	6	3.1	315	23	ABE76310	Human protein phos	334	6	3.1	402	22	ABW93220	Human protein sequ
262	6	3.1	315	23	ABP27294	Streptococcus poly	335	6	3.1	402	22	ABW5804	Human protein sequ
263	6	3.1	316	21	ABE56777	Human prostate can	336	6	3.1	402	23	ABW97296	Novel human protei
264	6	3.1	319	22	ABG06331	Novel human diagno	337	6	3.1	403	21	AG05758	Arabidopsis thalia
265	6	3.1	321	21	AG34423	Arabidopsis thalia	338	6	3.1	403	21	AG49835	Herbicidally activ
266	6	3.1	321	21	AAI94272	Corn arginyl-tRNA	339	6	3.1	404	23	ABW91257	Herbicidally activ
267	6	3.1	322	21	AAI92821	C. pneumoniae CPN1	340	6	3.1	406	20	AAI49152	Amino acid sequenc
268	6	3.1	323	22	AAW39441	Human polypeptide	341	6	3.1	406	21	AG05757	Arabidopsis thalia
269	6	3.1	324	22	ABW70087	Drosophila melanog	342	6	3.1	406	21	AG49834	Arabidopsis thalia
270	6	3.1	325	21	AG10582	Arabidopsis thalia	343	6	3.1	406	22	AG98360	Escherichia coli p
271	6	3.1	325	23	ABW47439	Listeria monocytog	344	6	3.1	408	21	AG37342	Arabidopsis thalia
272	6	3.1	326	21	AG10581	Arabidopsis thalia	345	6	3.1	408	21	AAI77123	Human neurotransmi
273	6	3.1	326	21	AG49782	Arabidopsis thalia	346	6	3.1	409	21	AG51364	Arabidopsis thalia
274	6	3.1	327	22	ABG04887	Novel human diagno	347	6	3.1	409	22	ABW93991	Human protein sequ
275	6	3.1	327	22	AG72811	Human olfactory re	348	6	3.1	410	17	AE91248	VIP2A(a) protein w
276	6	3.1	328	22	AAU51216	Propionibacterium	349	6	3.1	410	17	AAE91251	VIP2A(a) protein w
277	6	3.1	329	21	AG27834	Arabidopsis thalia	350	6	3.1	410	18	AAW19517	Maize optimised-B.
278	6	3.1	330	21	AG05509	Arabidopsis thalia	351	6	3.1	410	18	AAW19518	Maize optimised-B.
279	6	3.1	331	8	AAI70328	Biotin-synthase ge	352	6	3.1	410	19	AAW46728	Maize optimised VI
280	6	3.1	331	23	ABW11120	Herbicidally activ	353	6	3.1	410	19	AAW46729	VIP2A(a) with euka
281	6	3.1	332	21	AG29966	Arabidopsis thalia	354	6	3.1	412	20	AAI49151	Amino acid sequenc
282	6	3.1	335	23	ABW90580	Chlamydia pneumoni	355	6	3.1	413	20	AAI05724	Ras binding protei
283	6	3.1	336	22	ABW82592	S. epidermidis ope	356	6	3.1	413	22	ABW70347	Drosophila melanog
284	6	3.1	337	9	AAE81191	Biotin synthetase	357	6	3.1	415	22	AAU17462	Novel signal trans
285	6	3.1	339	21	AG10580	Arabidopsis thalia	358	6	3.1	417	23	AAW48190	Human keratin 45.8
286	6	3.1	339	21	AG49781	Arabidopsis thalia	359	6	3.1	419	20	AAI49243	N-terminal region
287	6	3.1	341	23	ABW7246	Mouse ischaemic co	360	6	3.1	419	20	AAI32182	N-terminal choline
288	6	3.1	342	21	AG34422	Arabidopsis thalia	361	6	3.1	419	20	AAI32182	Amino acid sequenc
289	6	3.1	346	21	AG32084	Arabidopsis thalia	362	6	3.1	420	21	AAI223364	Arabidopsis thalia
290	6	3.1	347	22	AAW41227	Human polypeptide	363	6	3.1	421	22	ABW64448	Drosophila melanog
291	6	3.1	352	21	AG29965	Arabidopsis thalia	364	6	3.1	422	21	AG05756	Arabidopsis thalia
292	6	3.1	352	23	ABW92713	Herbicidally activ	365	6	3.1	422	21	AG49833	Arabidopsis thalia
293	6	3.1	352	23	AAU76041	Human ADAM-H9 prot	366	6	3.1	425	20	AAI49248	N-terminal region
294	6	3.1	353	23	ABW49274	Listeria monocytog	367	6	3.1	425	20	AAI32187	N-terminal choline
295	6	3.1	354	20	AAI07104	Colon cancer assoc	368	6	3.1	425	21	AAI68783	Amino acid sequenc
296	6	3.1	354	21	AAE85290	Human G-alpha-11 a	369	6	3.1	425	22	AAW79121	Human protein SEQ
297	6	3.1	354	22	AAW99064	Human G-protein al	370	6	3.1	427	23	ABE41163	Human ovarian anti
298	6	3.1	354	23	ABW09273	G protein-coupled	371	6	3.1	430	21	AAI23363	Arabidopsis thalia
299	6	3.1	355	21	AAE29964	Arabidopsis thalia	372	6	3.1	431	21	AG32058	Arabidopsis thalia
300	6	3.1	355	21	AAE85149	Human G-alpha-12 a	373	6	3.1	432	22	ABW61904	Drosophila melanog
301	6	3.1	355	22	AAE99065	Human G-protein al	374	6	3.1	433	20	AAI49245	N-terminal region
302	6	3.1	355	23	ABW09274	G protein-coupled	375	6	3.1	433	20	AAI32184	N-terminal choline

84	6	3.1	69	21	AAG36767	Arabidopsis thalia	157	6	3.1	155	22	ABG27281	Novel human diagno
85	6	3.1	73	22	ABG24453	Novel human diagno	158	6	3.1	156	21	AAB57153	Human prostate can
86	6	3.1	73	22	AAU18639	Human lung antigen	159	6	3.1	156	22	ABG14998	Novel human diagno
87	6	3.1	73	22	AAB89495	Human immune/haema	160	6	3.1	157	21	AAB16272	Pinus radiata amyl
88	6	3.1	74	21	AAB16773	Bacteriophage Dp-1	161	6	3.1	157	21	AAB16272	Human protein sequ
89	6	3.1	76	21	AAB41435	Human ORFX ORF199	162	6	3.1	162	22	AAB95071	Streptococcus pneu
90	6	3.1	76	22	ABG05125	Novel human diagno	163	6	3.1	166	18	AAW14568	Arabidopsis thalia
91	6	3.1	78	21	AAU01097	Human secreted pro	164	6	3.1	166	22	AAB93936	Human protein sequ
92	6	3.1	79	22	AAG05346	Propionibacterium	165	6	3.1	173	21	AAG04363	Arabidopsis thalia
93	6	3.1	80	23	ABP3212	Human ORF2185 prot	166	6	3.1	173	23	AAU72498	Arabidopsis thalia
94	6	3.1	83	21	ABG54512	Zea mays protein f	167	6	3.1	179	20	AAU34845	Amino acid sequenc
95	6	3.1	85	22	AAB95735	Human testicular a	168	6	3.1	179	21	AAB25390	Pinus radiata cell
96	6	3.1	85	22	AAB95028	Human reproductive	169	6	3.1	180	23	ABW49350	Listeria monocytog
97	6	3.1	88	22	AAU22171	Human cardiovascu	170	6	3.1	184	18	AAW14589	Streptococcus pneu
98	6	3.1	89	22	AAO12560	Human polypeptide	171	6	3.1	186	23	ABG77118	Amino acid sequenc
99	6	3.1	90	21	AAU53798	Amino acids 145-23	172	6	3.1	193	22	ABG01292	Novel human diagno
100	6	3.1	96	22	AAU53798	Human immune/haema	173	6	3.1	196	19	AAW82630	Ehrlichia sp. HGE-
101	6	3.1	100	22	AAU53798	Human colon cancer	174	6	3.1	196	21	AAW82630	Ehrlichia sp. exte
102	6	3.1	98	21	AAB53527	Human endometrium	175	6	3.1	196	21	AAW78554	Ehrlichia antigen
103	6	3.1	105	20	AAU60007	Human endometrium	176	6	3.1	196	21	AAW78556	Ehrlichia antigen
104	6	3.1	105	21	AAG12168	Arabidopsis thalia	177	6	3.1	196	23	AAU73200	Human granulocytic
105	6	3.1	105	21	AAU90691	Human spectrin ple	178	6	3.1	196	23	AAU73213	Human granulocytic
106	6	3.1	105	22	AAU44839	Propionibacterium	179	6	3.1	198	22	ABG00682	Novel human diagno
107	6	3.1	107	23	ABW90676	Chlamydia pneumoni	180	6	3.1	198	22	ABG67520	Amino acid sequenc
108	6	3.1	107	23	AAU48199	Human zinc finger	181	6	3.1	198	23	ABP38614	Staphylococcus epi
109	6	3.1	109	22	ABG30376	Novel human diagno	182	6	3.1	204	22	ABW70670	Drosophila melanog
110	6	3.1	110	23	ABP30180	Streptococcus poly	183	6	3.1	206	22	ABW69257	Drosophila melanog
111	6	3.1	111	20	AAU37080	Chlamydia trachoma	184	6	3.1	212	18	AAW14588	Streptococcus pneu
112	6	3.1	111	23	ABP42966	Human ovarian anti	185	6	3.1	214	20	AAU35473	Chlamydia pneumoni
113	6	3.1	113	22	AAU66759	Propionibacterium	186	6	3.1	219	21	AAU49000	Arabidopsis thalia
114	6	3.1	113	22	ABG24902	Novel human diagno	187	6	3.1	220	23	ABP01204	Human ORFX protein
115	6	3.1	117	23	ABP26893	Streptococcus poly	188	6	3.1	221	21	AAB41753	Human ORFX ORF1517
116	6	3.1	118	22	AAU00051	Human protein sequ	189	6	3.1	221	21	AAU94283	Corn arginyl-tRNA
117	6	3.1	118	22	AAU02054	Synthetic human ta	190	6	3.1	223	21	AAG48999	Arabidopsis thalia
118	6	3.1	119	18	AAU23002	Canine herpesvirus	191	6	3.1	224	22	AAU94724	Human protein sequ
119	6	3.1	119	19	AAU72662	Canine herpes viru	192	6	3.1	227	20	AAU10844	Amino acid sequenc
120	6	3.1	120	22	AAU40633	Canine herpes viru	193	6	3.1	228	21	AAU49931	Arabidopsis thalia
121	6	3.1	126	22	ABG67055	Propionibacterium	194	6	3.1	230	23	ABW54441	Lactococcus lactis
122	6	3.1	126	23	ABG67055	Drosophila melanog	195	6	3.1	232	21	AAG49930	Arabidopsis thalia
123	6	3.1	127	23	ABW7947	Listeria monocytog	196	6	3.1	232	18	AAW14590	Streptococcus pneu
124	6	3.1	130	22	ABG23863	Novel human diagno	197	6	3.1	235	22	AAG90926	C glutamicum prote
125	6	3.1	132	21	AAU41950	Human ORFX ORF1714	198	6	3.1	236	22	AAG92550	C glutamicum prote
126	6	3.1	132	23	ABW55438	Lactococcus lactis	199	6	3.1	238	18	AAW14587	Streptococcus pneu
127	6	3.1	134	21	ABW58888	Breast and ovarian	200	6	3.1	238	22	ABG15228	Novel human diagno
128	6	3.1	134	21	AAG55969	Zea mays protein f	201	6	3.1	240	21	AAU94275	Wheat arginyl-tRNA
129	6	3.1	136	22	ABW92870	Human protein sequ	202	6	3.1	242	22	AAU79067	Human protein sequ
130	6	3.1	138	22	ABW27701	Human peptide #352	203	6	3.1	242	22	AAU39926	Human polypeptide
131	6	3.1	138	22	ABW32872	Peptide #378 encod	204	6	3.1	243	15	AAW63792	Bacillus cereus ve
132	6	3.1	138	22	ABW18354	Protein #353 encod	205	6	3.1	243	23	ABW47402	Listeria monocytog
133	6	3.1	138	22	AAW53674	Human brain expres	206	6	3.1	244	23	ABP39939	Staphylococcus epi
134	6	3.1	138	22	AAW66058	Human bone marrow	207	6	3.1	247	21	AAU74389	Neisseria gonorrhe
135	6	3.1	138	22	AAU13927	Peptide #361 encod	208	6	3.1	254	23	ABW79353	Human ovary specif
136	6	3.1	138	22	AAU26333	Peptide #370 encod	209	6	3.1	254	23	ABW77351	Plodia interpuncte
137	6	3.1	138	22	AAU01670	Peptide #352 encod	210	6	3.1	255	22	AAU94259	Human protein sequ
138	6	3.1	138	23	ABW35706	Human peptide enco	211	6	3.1	258	21	AAU56510	Human Jurkat cell
139	6	3.1	139	18	AAU23003	Canine herpesvirus	212	6	3.1	259	22	AAU39028	Human polypeptide
140	6	3.1	139	19	AAU72663	Canine herpes viru	213	6	3.1	259	23	ABW97275	Novel human protei
141	6	3.1	139	22	AAU51320	Canine herpes viru	214	6	3.1	261	20	AAU37150	Chlamydia trachoma
142	6	3.1	140	21	AAB31354	Human ORFX ORF2918	215	6	3.1	261	22	AAW14712	Human polypeptide
143	6	3.1	141	22	AAU41819	Human polypeptide	216	6	3.1	261	22	AAU93180	Human protein sequ
144	6	3.1	142	21	AAG04365	Arabidopsis thalia	217	6	3.1	261	23	ABW53449	Lactococcus lactis
145	6	3.1	143	19	AAU96163	S. pneumoniae deri	218	6	3.1	262	22	ABW64085	Drosophila melanog
146	6	3.1	145	22	AAU87315	Novel central nerv	219	6	3.1	263	21	AAG11047	Arabidopsis thalia
147	6	3.1	146	22	AAU16613	Human novel secret	220	6	3.1	263	21	AAG40402	Arabidopsis thalia
148	6	3.1	147	22	ABG04065	Novel human diagno	221	6	3.1	265	21	AAU94451	Human inflammation
149	6	3.1	148	21	AAG04364	Arabidopsis thalia	222	6	3.1	266	21	AAU11046	Arabidopsis thalia
150	6	3.1	148	21	AAG14968	Arabidopsis thalia	223	6	3.1	266	21	AAG40401	Arabidopsis thalia
151	6	3.1	148	22	AAB95671	Putative P. abyssal	224	6	3.1	267	22	AAG40814	Human polypeptide
152	6	3.1	148	22	AAB66021	Human secreted pro	225	6	3.1	271	22	AAW41266	Human polypeptide
153	6	3.1	149	18	AAW22993	Canine herpesvirus	226	6	3.1	273	21	AAG27836	Arabidopsis thalia
154	6	3.1	149	19	AAU72652	Canine herpes viru	227	6	3.1	276	21	AAG27835	Arabidopsis thalia
155	6	3.1	149	20	AAU34841	Chlamydia pneumoni	228	6	3.1	276	21	AAG31086	Arabidopsis thalia
156	6	3.1	149	22	AAU51309	Canine herpes viru	229	6	3.1	279	21	AAG32062	Arabidopsis thalia

GenCore version 5.1.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2003, 09:54:48 : Search time 36 Seconds
(without alignments)
710.670 Million cell updates/sec

Title: US-09-868-352-23

Perfect score: 192

Sequence: 1 MRKEVTEPLNKNYGPQF.....EEQDKEMTSKQHLFLVRKN 192

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

A_Geneseq_101002:*

1:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20:	/SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21:	/SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22:	/SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23:	/SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	192	100.0	192	21	AA196805
2	185	96.4	192	23	ABP25486
3	12	6.2	176	22	AAU37821
4	12	6.2	22	AAU37974	Streptococcus pneu
5	12	6.2	214	23	ABP25487
6	11	5.7	91	23	ABP25067
7	8	4.2	251	23	ABP53482
8	8	4.2	303	23	ABP28656
9	8	4.2	733	20	AA130937
10	8	4.2	733	20	AA122202

11	7	3.6	15	20	AAW93973	Human IRS-1 and IR
12	7	3.6	129	22	AG82114	S. epidermidis ope
13	7	3.6	132	23	ABP40049	Staphylococcus epi
14	7	3.6	157	22	ABP47229	Yeast Mfn. Saccha
15	7	3.6	159	20	AAW93976	Human IRS-1 and IR
16	7	3.6	235	22	AAW40165	Human polypeptide
17	7	3.6	240	21	AAW09871	Arabidopsis thalia
18	7	3.6	248	21	AAW09870	Arabidopsis thalia
19	7	3.6	254	22	ABP36598	Human FLEXHT-20 pr
20	7	3.6	256	21	AAW09869	Arabidopsis thalia
21	7	3.6	266	22	AAU30761	Novel human secret
22	7	3.6	280	21	AAW42139	Human ORFX ORF1903
23	7	3.6	283	22	AAW41951	Human polypeptide
24	7	3.6	284	20	AAW41211	H. influenzae Vig
25	7	3.6	317	19	AAW73507	Human ATG-1709 pro
26	7	3.6	317	19	AAW73507	Human secreted apo
27	7	3.6	324	22	AAU54873	Propionibacterium
28	7	3.6	334	21	AAW41766	Human ORFX ORF1530
29	7	3.6	376	22	AAW51364	Feline infectious
30	7	3.6	380	9	AAW82001	Sequence of phenyl
31	7	3.6	382	14	AAW39224	Nucleocapsid prote
32	7	3.6	458	23	ABW77059	Human protein sequ
33	7	3.6	628	23	ABW54651	Lactococcus lactis
34	7	3.6	782	23	AAU81501	Chlamydomonas inte
35	7	3.6	823	21	AAW42436	Arabidopsis thalia
36	7	3.6	849	21	AAW42816	Human ORFX ORF2580
37	7	3.6	1242	20	AAW13461	Amino acid sequenc
38	7	3.6	1242	20	AAW93972	Human IRS-1 and IR
39	7	3.6	1242	22	AAW83921	Amino acid sequenc
40	7	3.6	1243	16	AAW67708	Insulin receptor s
41	7	3.6	1321	17	AAW69994	Mouse IRS-2. Mus
42	7	3.6	1489	22	ABG19124	Novel human diagno
43	7	3.6	2386	18	AAW13153	S. pombe Rad3 poly
44	6	3.1	19	21	AAW98312	Alpha D peptide de
45	6	3.1	25	22	AAW30370	Novel human secret
46	6	3.1	29	22	ABG29861	Novel human diagno
47	6	3.1	38	15	AAW50672	G-protein coupled
48	6	3.1	38	17	AAW02864	G-protein coupled
49	6	3.1	38	22	ABW40724	Peptide #8230 enco
50	6	3.1	38	22	ABW24951	Protein #6950 enco
51	6	3.1	38	22	AAW61586	Human brain expres
52	6	3.1	38	22	AAW74377	Human bone marrow
53	6	3.1	38	22	AAW20290	Peptide #6724 enco
54	6	3.1	38	22	AAW34490	Peptide #8527 enco
55	6	3.1	38	23	ABG44284	Human peptide enco
56	6	3.1	39	22	ABG18021	Novel human diagno
57	6	3.1	42	22	ABW31406	Peptide #4057 enco
58	6	3.1	42	22	ABW36620	Peptide #4126 enco
59	6	3.1	42	22	ABW21956	Protein #3955 enco
60	6	3.1	42	22	AAW57381	Human brain expres
61	6	3.1	42	22	AAW69776	Human bone marrow
62	6	3.1	42	22	AAW17600	Peptide #4034 enco
63	6	3.1	42	22	AAW30121	Peptide #4158 enco
64	6	3.1	42	22	AAW05261	Peptide #3943 enco
65	6	3.1	42	23	ABG39406	Human peptide enco
66	6	3.1	48	22	ABG03951	Novel human diagno
67	6	3.1	49	22	ABW29257	Peptide #1908 enco
68	6	3.1	49	22	ABW34422	Peptide #1928 enco
69	6	3.1	49	22	ABW19834	Protein #1833 enco
70	6	3.1	49	22	AAW55212	Human brain expres
71	6	3.1	49	22	AAW67607	Human bone marrow
72	6	3.1	49	22	AAW15414	Peptide #1848 enco
73	6	3.1	49	22	AAW27903	Peptide #1940 enco
74	6	3.1	49	22	AAW03176	Peptide #1858 enco
75	6	3.1	49	23	ABG37197	Human peptide enco
76	6	3.1	51	22	AAU45186	Propionibacterium
77	6	3.1	53	21	AAW53800	Amino acids 145-19
78	6	3.1	54	23	ABP07204	Human ORFX protein
79	6	3.1	56	22	AAW90475	Human immune/haema
80	6	3.1	64	22	AAU55056	Propionibacterium
81	6	3.1	66	21	AAW91503	Human secreted pro
82	6	3.1	67	22	ABG15925	Novel human diagno
83	6	3.1	69	21	AAW12917	Arabidopsis thalia

```
QY 130 LysSerArgSerLys---SerGlnSerSerLysSerGln-----ThrArgAsnAsn 145
Db ::::||||| ||| ||||||| ::::||||| ||| |||||||
311 CGAACCCTCGAANAACGTCCCAACGACAAACAAACACGACCCGACCCGAAACTAC 370
QY 146 ArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysLysArgLysAspThrLysArg 165
Db ::::||||| ||| ||||||| ::::||||| ||| |||||||
371 AACATCAAAACGCGCTAAACACTAAACGCGACACCAACCAACGCAACGCGTAACCGA 430
QY 166 ArgGlnGluArgHisTle-----LysGluGluGlnAspLys-GluMetTh 180
Db ||| ::::||||| ||| ||||||| ::::||||| ||| |||
431 CGACTTAACCTCGAATTACAAAAACCGAAGAACTCCGCGAAAAAATAACTCAAC 490
QY 180 rSerAlaLysGlnHisLeuLeuPheValArgLysAsn 192
Db ||| ::::||||| ||| |||
491 CTGGAACCGACACATCTCTCTCCCTCCCGCGCCTAAC 527
```

Search completed: February 25, 2003, 10:59:34
Job time : 1267 secs

```
PD 07-MAR-2002.
XX ABQ50645 standard; DNA; 528 BP.
XX AC ABQ50645;
XX DT 12-JUL-2002 (first entry)
XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 37236.
XX KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX KW drug; side effect; cancer; central nervous system; cardiovascular;
XX KW gastrointestinal; respiratory system; single nucleotide polymorphism;
XX KW SNP; cell differentiation; ds.
XX OS Homo sapiens.
XX PN WO200218632-A2.
XX PD 07-MAR-2002.
XX PF 01-SEP-2001; 2001WO-EP10074.
XX PR 01-SEP-2000; 2000DE-1043826.
XX PR 05-SEP-2000; 2000DE-1044543.
XX PA (EPIG-) EPIGENOMICS AG.
XX PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX WPI; 2002-371829/40.
XX DR Determining the degree of cytosine methylation in genomic DNA, useful
XX PT for diagnosis and prognosis, comprises selective hybridization of
XX PT amplicons from chemically treated DNA
XX PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX CC This invention describes a novel method for determining the degree of
XX CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
XX CC genomic sample of DNA. The sample is treated chemically to convert
XX CC cytosine (C) but not methylated C, to uracil, then part of the genomic
XX CC DNA that contains the target C is amplified to form a labeled amplicon.
XX CC The amplicon is hybridised to two classes, each with at least one
XX CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
XX CC and the degree of hybridisation to both classes is determined from the
XX CC label on the amplicon. From the ratio of labels hybridised to the two
XX CC classes of oligomers, the degree of methylation is calculated. The method
XX CC is used: (i) for diagnosis and/or prognosis of side effects of
XX CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
XX CC of the central nervous, cardiovascular, gastrointestinal and respiratory
XX CC systems etc., particularly by detecting mutations or single nucleotide
XX CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
XX CC types and for investigating cell differentiation. The method allows the
XX CC methylation status of many C residues to be determined simultaneously.
XX CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
XX CC method for determining the degree of cytosine methylation described in
XX CC the disclosure of the invention.
XX SQ Sequence 528 BP; 64 A; 55 C; 177 G; 232 T; 0 other;

Alignment Scores:
Pred. No.: 0.0548 Length: 528
Score: 98.50 Matches: 34
Percent Similarity: 49.46% Conservative: 12
Best Local Similarity: 36.56% Mismatches: 37
Query Match: 9.76% Indels: 10
DB: 24 Gaps: 3

US-09-868-352-23 (1-192) x ABQ50644 (1-528)
Qy 110 LeuGluAsnProAsnProArgAspIleLysPheAspAspGluArgProHisLysArgArg 129
Db 278 ATACCCGACCGGACCCACCTCCCGAAATCAACAAACAACTCGAACACGAAACGAAAA 219
Qy 130 LysSerArgSerLys---SerGlnSerSerLysSerGln-----ThrArgAsnAsn 145
Db 218 CGAACCGTCGAAAACGCTCCACGACACAAACAAACAAACGACCCGACCCGAACTAC 159
Qy 146 ArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysLysArgLysAspThrLysArg 165
Db 158 AACATCAAAACGCGCTAAACACTAAACGACACACAAACAAACAAACGACCGTAAACCGA 99
Qy 166 ArgGlnGluArgHisIle-----LysGluGluGlnAspLys-GluMetTh 180
Db 98 CGACTTAACCTCGAAATTCACAAAACAAACGACGAAACCTCCGCGAAAAATAACTCAAC 39
Qy 180 rSerAlaLysGlnHisLeuPheValArgLysAsn 192
Db 38 CTCGAACCGACACACATCTCTCCCTCCCGCCCTTAC 2

RESULT 40
```

CC sequence represents a gene differentially expressed in granulocytes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2736 BP; 867 A; 521 C; 608 G; 740 T; 0 other;

Alignment Scores:
Pred. No.: 0.422 Length: 2736
Score: 99.00 Matches: 24
Percent Similarity: 58.57% Conservative: 17
Best Local Similarity: 34.29% Mismatches: 21
Query Match: 9.81% Indels: 8
DB: 24 Gaps: 2

US-09-868-352-23 (1-192) x ABA84078 (1-2736)
QY 122 AspGluArgProHisLysArgArgLysSerArgSerLysSerGlnSer----- 137
DB 801 GATAAGAAAGAGAAAGAGGCGATTCAAGATCAAGATCAAGTCTTAGGAGGAGGAGG 860
QY 138 -----SerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAlaHisPhe 155
DB 861 ACTCCCTCATCTCTAGACACAGCGCGTCAAGAGGAGATCGAGCGGCGTCAATTCT 920
QY 156 ThrSerLysLysArgLysAspThrLys-----ArgArgGlnGluArgHisIleLysGlu 173
DB 921 AAGTCTAGGAGTGGCGGACGATCCAAAGCCCAAGCGGAGAGATCTCATTCACAGAA 980
QY 174 GluGlnAspLysGluMetThrSerAlaLys 183
DB 981 AGAGGTAGAGGTCAAGGACACATCAAAA 1010

RESULT 38
AAC78163
ID AAC78163 standard; cDNA; 3484 BP.
XX
AC AAC78163;
XX
XX
DT 08-FEB-2001 (first entry)
XX
DE Human cancer associated gene sequence SEQ ID NO:557.
XX
KW Human; cancer associated gene; cancer antigen; detection; cancer;
KW diagnosis; cytostatic; proliferative; vulnary; immunomodulator;
KW antidiabetic; antisthmatic; antirheumatic; antithyroid; antiviral;
KW antiinflammatory; antithyroid; antiallergic; antibacterial; cardiant;
KW dermatological; neuroprotective; thrombolytic; coagulant; neutropic;
KW vasotropic; antipsoriatic; antidiabetic; antidiabetic; antidiabetic;
KW immune disorder; haematopoietic cell disorder; gene therapy; inflammation;
KW allergic reaction; graft versus host disease; organ rejection;
KW haemostatic; thrombolytic; cardiovascular disorder; infection;
KW neurological disease; drug screening; ss.
XX
OS Homo sapiens.
XX
XX WO200055350-A1.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US05882.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
XX WPI; 2000-587533/55.
XX P-PSDB; ABA43954.
XX
XX Novel isolated nucleic acids comprising sequences encoding peptides

PT useful for treating or diagnosing e.g. cancer -
XX
PS Claim 1; Page 1083-1084; 2352pp; English.
XX
CC AAC77607 to AAC78448 encode the human cancer associated proteins given
CC in ABA43398 to ABA44239. The proteins can have activities based on the
CC tissues and cells the genes are expressed in. Example of activities
CC include: cytostatic; proliferative; vulnary; immunomodulator;
CC antidiabetic; antisthmatic; antirheumatic; antithyroid; antiviral;
CC antiinflammatory; antithyroid; antiallergic; antibacterial; cardiant;
CC dermatological; neuroprotective; cardiant; thrombolytic; coagulant;
CC neutropic; vasotropic; antipsoriatic and antidiabetic. The
CC polynucleotides and polypeptides can be used for preventing, treating or
CC ameliorating medical conditions and diagnosing pathological conditions.
CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
CC the present invention may be used to treat immune disorders by activating
CC or inhibiting the proliferation, differentiation or mobilisation of
CC immune cells, to treat disorders of haematopoietic cells, autoimmune
CC disorders, allergic reactions, graft versus host disease and organ
CC rejection, modulate haemostatic or thrombolytic activity, modulate
CC inflammation, cancers, cardiovascular disorders, neurological disease and
CC bacterial or viral infections. The peptides, nucleotides, antibodies,
CC agonists and antagonists may be also be used in drug screens. AAC78449 to
CC AAC78457 and ABA44240 represent sequences used in the exemplification of
CC the present invention.
XX
SQ Sequence 3484 BP; 1111 A; 607 C; 695 G; 1068 T; 3 other;

Alignment Scores:
Pred. No.: 0.58 Length: 3484
Score: 99.00 Matches: 24
Percent Similarity: 58.57% Conservative: 17
Best Local Similarity: 34.29% Mismatches: 21
Query Match: 9.81% Indels: 8
DB: 21 Gaps: 2

US-09-868-352-23 (1-192) x AAC78163 (1-3484)
QY 122 AspGluArgProHisLysArgArgLysSerArgSerLysSerGlnSer----- 137
DB 1101 GATAAGAAAGAGAAAGAGGCGATTCAAGATCAAGATCAAGTCTTAGGAGGAGGAGG 1160
QY 138 -----SerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAlaHisPhe 155
DB 1161 ACTCCCTCATCTCTAGACACAGCGCGTCAAGAGGAGATCGAGCGGCGTCAATTCT 1220
QY 156 ThrSerLysLysArgLysAspThrLys-----ArgArgGlnGluArgHisIleLysGlu 173
DB 1221 AAGTCTAGGAGTGGCGGACGATCCAAAGCCCAAGCGGAGAGATCTCATTCACAGAA 1280
QY 174 GluGlnAspLysGluMetThrSerAlaLys 183
DB 1281 AGAGGTAGAGGTCAAGGACACATCAAAA 1310

RESULT 39
ABQ50644/c
ID ABQ50644 standard; DNA; 528 BP.
XX
AC ABQ50644;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 37235.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
XX WO200218632-A2.
XX

CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
CC reperfusion injury, ARDS, adult respiratory distress syndrome,
CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,
CC periodontal disease; also bacterial infection, viral infection,
CC parasitic infection, protozoal infection, fungal infection and M5 is
CC useful for treating one of the above conditions. The present
CC sequence represents a gene differentially expressed in granulocytes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2076 BP; 597 A; 467 C; 611 G; 401 T; 0 other;

US-09-868-352-23 (1-192) x ABK83839 (1-2076)

Qy	17	GlyProGlnPheIleHisPheGluAsnIleValLysSerAspAspIleGluPheGlnLeu	36
Db	153	GGATATGCTTTGCGAGTTTGATGCTCGTGATCGACATGATGCTGTTTATGAACCT	212
Qy	37	ValIleAsnGluLysSerAlaPheAspValThrVal-	48
Db	213	ANTGGCAACACCTTGTGTGGTACCGAGTAATGTTGACGATCCCGCGGCCACGGCGA	272
Qy	49	-----PheGlnArgPheSer	54
Db	273	GATGGCAGTTACGGTCTCGACGCAGTGGATATGCTTATAGAAGAAGTGGCCGAGATAAA	332
Qy	55	-----GluIleLeuLeuLysTyrAspPheIleValClyAsp-	66
Db	333	TATGGCCCTCCTACTCGCACAGATACAGACTTATTGTGGAAATTTGTCAAGTCGGTGC	392
Qy	67	---TrpGlyAsn-----GluGlnLeuArgLeuArgGly-	81
Db	393	AGCTTGGCAAGACCTAAAGGATTATATCGCTCAGGCAGGAGAGTAGTACTTATGCAGATGCT	452
Qy	82	SerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyrCys	101
Db	453	CACAAGGAGCGCAAAAT-----GAAGGGGTGATTGAATTTGTA	491
Qy	102	AsnPheGlyCysAlaTyrPheValLeuGlu-	116
Db	492	TCATTATTCTGATATGAAGAGAGCTTTGGAAAGATTTGATGGAACCTCAAGTGGGAGA	551
Qy	117	AspIleLysPheAspAspGluArgPro-	125
Db	552	AAAATCAGATTAGTTGAAGACAAGCCAGGGTCCAGACGACGCGGTCCTACTCCAGAAGC	611
Qy	126	-----HisLysArg	135
Db	612	CGGAGTCATTCAAGGTCTCGTCTCGAAGCAGACATTCCTCCGTAAGAGCAGAGCCCAAGT	671
Qy	136	GlnSerSerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPhe	155
Db	672	GGCAGCAGCAAAAGCAGCTATTCTAAGAGTAGATCTCGTCCAGGTCGGGCTCCCGCTCC	731
Qy	156	ThrSerLys-----LysArgLysAspThrLysArgArgGlnGluArgHisIleLysGlu	173
Db	732	CGGAGCAGAGCCGGACCGGACCCAGATGTCGGAGCCGAGCAGAAAGAGAAAACGAGG	791
Qy	174	GluGlnAspLysGluMetThrSerAlaLysGlnHis	185
Db	792	AGCCCCAGCAAGGACAAGAGCCGACCGCAGCCAT	827

KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
 KW adult respiratory distress syndrome; inflammatory bowel disease;
 KW Crohn's disease; ulcerative colitis; periodontal disease;
 KW granulocyte activation; chronic inflammation; allergy.
 XX
 OS Homo sapiens.
 XX
 XX WO200228999-A2.
 XX
 XX 11-APR-2002.
 XX
 XX 03-OCT-2001; 2001WO-US30821.
 XX
 XX 03-OCT-2000; 2000US-237189P.
 XX
 XX (GENE-) GENE LOGIC INC.
 XX
 XX Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
 XX WPI; 2002-435328/46.
 DR
 XX
 XX Detecting granulocyte activation by detecting differential expression
 PT of genes associated with granulocyte activation, which serves as
 PT diagnostic markers that is useful for monitoring disease states and
 PT drug toxicity -
 XX
 XX Claim 1; SEQ ID No 859; 114pp; English.
 XX
 CC The invention relates to detecting (M1) granulocyte (GC) activation
 CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
 CC DNA chip analysis as given in the specification, and comparing
 CC the expression level to an expression level in an unactivated
 CC GC, where differential expression of Gs is indicative of GCA.
 CC Also included are modulating (M2) GA by contacting GC with an agent
 CC that alters the expression of at least one gene in Gs; (2) screening (M3)
 CC for an agent capable of modulating GCA or an inflammation (especially
 CC chronic) in a tissue, an allergic response in a subject, exposure of a
 CC subject to a pathogen or sterile inflammatory disease using the
 CC gene expression profile; (3) detecting (M4) an inflammation (especially
 CC chronic) in a tissue, an allergic response in a subject, exposure of a
 CC subject to a pathogen or sterile inflammatory disease, by detecting the
 CC level of expression in a sample of the tissue of gene(s) from Gs, where
 CC the level of expression of the gene is indicative of inflammation;
 CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,
 CC an allergic response in a subject, exposure of a subject to a pathogen
 CC or sterile inflammatory disease, by contacting a tissue having
 CC inflammation with an agent that modulates the expression of gene(s)
 CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
 CC modulating GA; M3 is useful for screening an agent capable of modulating
 CC GCA preferably in an inflammation in a tissue; M4 is useful for
 CC detecting an inflammation (especially chronic) in a tissue, an allergic
 CC response in a subject, exposure of a subject to a pathogen or sterile
 CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
 CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
 CC reperfusion injury, ARDS, adult respiratory distress syndrome,
 CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,
 CC periodontal disease; also bacterial infection, viral infection,
 CC parasitic infection, protozoal infection, fungal infection and M5 is
 CC useful for treating one of the above conditions. The present
 CC sequence represents a gene differentially expressed in granulocytes.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1454 BP; 432 A; 255 C; 365 G; 402 T; 0 other;

Alignment Scores:

Pred. No.:	0.161	Length:	1454
Score:	99.50	Matches:	43
Percent Similarity:	41.23%	Conservative:	44
Best Local Similarity:	20.38%	Mismatches:	73
Query Match:	9.86%	Indels:	51

DB:	24	Gaps:	8
US-09-868-352-23 (1-192) x ABK84288 (1-1454)			
Qy	20	PheIleHisPheGluAsnIleValLysSerAspIleGluPheGlnLeuValIleAsn	39
Db	210	TTTGTGGAAATTTGAGGATCCAAAGGATCGAGATGATGCTGTGTATGAGCTTGTGGA	269
Qy	40	GlulysSerAlaPheAspValThrVal-----	48
Db	270	GAACTCTGTAGTGAAGGGTTACTATTGAACATGCTAGGGCTCGGTACAGAGGTGAAGA	329
Qy	49	-----PheGlyGlnArgPheSerGluIleLeuLysTyrAsp-----	61
Db	330	GGTAGAGGACGATACTCTGACCGCTTTTAGTAGTCGCACACCTCGAAATGATAGACGAAT	389
Qy	62	-----PheIleValGlyAsp-----	66
Db	390	GCTCCACCTGTAGAACACAGAAAATCGTCTTATAGTTGAGAATTTATCTCAAGATCAGC	449
Qy	67	TrpGlyAsnGluGlnLeuArgGlyPheTyrLysAspAlaSerThrIleArg---	85
Db	450	TGGCAGGAT-----CTCAAGATTTCATGACACAGCTGGGGAAAGTAAACGTTT	497
Qy	86	LysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyrCysAsnPheGlyCys	105
Db	498	CGGATGTCACACCGACCTAAATTAATGAAGGGTGGTTGAGTTTGCTCTTATGGTGAC	557
Qy	106	AlaTyrPheValLeuGlu-----AsnProAsnProArgAspIleLys---	119
Db	558	TTAAGAATGCTATTGAAATAAATCTTGGAAAGGAATAAATGGAGAAAAATAAAATTA	617
Qy	120	PheAspAspGluArgProHisLysArgLysSerArgSerLysSerGlnSer-----	137
Db	618	ATTGAAGGCAGCAAGAGCAGATAGGTCAAGAAGCAGGTCTCGATCCCGACCAAGAGT	677
Qy	138	-----SerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAlaAsn	152
Db	678	TCCTCTAGGTCTCTAGCCGATCCGTCCTCGTAGTCCCAATCTTACAGCCGGTCAAGA	737
Qy	153	AlaHisPheThrSerLysLysArgLysAspThrLysArgGlnGluArgHisIleLys	172
Db	738	AGCAGGACGAGGAGCGGAGCGGAGCAAGTCCCGTCTCTGTAGTAGTCTCCCGTGCCT	797
Qy	173	GluGluGlnAspLysGluMetThrSerAlaLys	183
Db	798	GAGAAAGAGCCAGAAACGTTGTTCTTCAAGTAGA	830
RESULT	35		
AAC75975			
ID	AAC75975	standard; cDNA; 1411 BP.	
XX	AAC75975;		
XX			
XX	08-FEB-2001	(first entry)	
XX	Human ORF1530	polynucleotide sequence SEQ ID NO:3059.	
XX	Human; open reading frame; ORF; detection; cytostatic; hepatotropic;		
XX	vulnerable; antipsoriatic; antiparkinsonian; neurotropic; neuroprotective;		
XX	anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;		
XX	immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;		
XX	hypotensive; dermatological; immunosuppressive; antiinflammatory;		
XX	antiviral; antibacterial; antifungal; antirheumatic; antithyroid;		
XX	antianemic; gene therapy; cancer; proliferative disorder; hypertension;		
XX	neurodegenerative disorder; osteoarthritis; graft vs host disease;		
XX	cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;		
XX	cholesterol ester storage; systemic lupus erythematosus; infection;		
XX	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;		
XX	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;		
XX	bone damage; cartilage damage; antiinflammatory disease; coagulation;		
XX	thrombosis; contraceptive; ss.		

US-09-868-352-23 (1-192) x AB035135 (1-886)

QY 112 AsnProAsnProArgAspIleLysPheAspGluArgProHisLysArgArgLysSer 131
 DB 448 AACACGAATCCACGCTACTCGACGTATCGTCTCGTCCCAACGCTAGGACTAATCA 507
 QY 132 ArgSerLysSerGlnSerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAla 151
 DB 508 ACGCGCTCTCCAATAACGCGCGCGACACACGCGACAAACGCGTCTCGCTTAACCC 567
 QY 152 AsnAlaHisPheThrSerLysLysArgLysAspThrLysArgArgGlnGluArgHisIle 171
 DB 568 ATAATGACCTCTCTACCAAAAGAAACGACCACTCGTCTGACACACGTACAAAT--- 624

QY 172 LysGluGluGlnAspLysGluMetThrSerAlaLysGlnHis 185

DB 625 ---AACCAATACGATAAAATAACAACAACCTCAAAAAAACA 663

RESULT 30

AAAT70121

ID AAA70121 standard; DNA; 1527 BP.

AC AAA70121;

XX 07-NOV-2000 (first entry)

DT Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:254.

DE Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
 KW antimalarial; malaria; protozoa; infection; insecticide; ds.
 KW Plasmodium falciparum.

OS WO200025728-A2.

PN 11-MAY-2000.

PD 05-NOV-1999; 99WO-0526796.

XX 05-NOV-1998; 98US-0107131.

XX (HOFF/) HOFFMAN S.

PA (CARU/) CARUCCI D.

PA (GARD/) GARDNER M.

PA (VENT/) VENTER J C.

XX Hoffman S, Carucci D, Gardner M, Venter JC;

PI WPI; 2000-365347/31.

DR Proteins encoded by chromosome 2 of the human malarial parasite,

XX Plasmodium falciparum, useful as antimalarial vaccines and in the

PT diagnosis of P.falciparum infection.

PS Disclosure; Page 472; 577pp; English.

XX

CC The present invention describes proteins and their fragments (I) encoded

CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.

CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)

CC vaccines against P. falciparum infection comprising (I) or (II).

CC (I) and (II) are useful for the development of vaccines against

CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal

CC antibody raised to immunogens comprising the sequences of (I), are

CC useful in the detection of infection with P. falciparum. Furthermore,

CC (I) (especially when they are rifins or secreted or membrane proteins)

CC can aid the identification of drugs to treat or prevent P. falciparum

CC infection, or they can be used to identify drug resistance in

CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the

CC subsequent identification of proteins encoded by it will help to expand

CC our understanding of parasite biology, a process hampered by the

CC complexity of the parasitic lifecycle, and provide new targets for

CC vaccine and drug development. Parasite resistance to drugs and mosquito

CC resistance to insecticides have led to a resurgence of malaria in many

CC parts of the world, and there is a pressing need for vaccines and new
 CC drugs. AAAT70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
 CC and protein sequences given in the present invention, but which are not
 CC specifically mentioned within the specification.

SQ Sequence 1527 BP; 848 A; 94 C; 210 G; 375 T; 0 other;

Alignment Scores:

Pred. No.: 0.116 Length: 1527
 Score: 101.00 Matches: 41
 Percent Similarity: 40.31% Conservative: 38
 Best Local Similarity: 20.92% Mismatches: 61
 Query Match: 10.01% Indels: 56
 DB: 21 Gaps: 9

US-09-868-352-23 (1-192) x AAA70121 (1-1527)

QY 5 ValThrProGluMetLeuAsnTyrAsnLysTyrProGlyProGlnPheIleHisPheGlu 24

DB 790 ATAACAAAAGAA---TGTAATGTTAATAATATGAT-----GAT 825

QY 25 AsnIleVal-----LysSerAspAspIleGlu--- 33

DB 826 AATATTATAGATCACAACAAACACAGGAAACAAAGAAAGAAAGAAAGTATCGAAAT 885

QY 34 -----PheGlnLeuValIleAsnGluLysSerAlaPheAspValThrValPheGly 50

DB 886 ATCAATATAAGCCATATATATGAAAGAAACAATCTCATGACATTTGT----- 936

QY 51 GlnArgPheSerGluIleLeuLysTyrAspPheIleValGlyAspTrpGlyAsnGlu 70

DB 937 -----AATGTATTAGAAGAAATAAAGAG 960

QY 71 GlnLeuArgLeuArgGlyPheTyrLysAspAlaSerThrIleArgLysAsnSerArgIle 90

DB 961 GAACAAAATATAACAATTTACAAAAGAGTCTTATACAAATTTGTAATATGTAAGGTA 1020

QY 91 SerArgLeuGluAspTyr-----IleLysGluTyrCysAsnPheGlyCysAlaTyrPhe 108

DB 1021 ---AAACTTGAGAAATATCATCATGAAAGAAATTTGAATAAT----- 1059

QY 109 ValLeuGluAsnProAsnProArgAspIleLysPheAspAspGluArgProHisLysArg 128

DB 1060 ---GTTCAAAATTAATAATGATGATGATTTAAAAGAAAGATGAGGCAAGAGAAAGAAAT 1116

QY 129 ArgLysSerArgSerLysSerGlnSerLysSerGlnThrArgAsnAsnArgSerGln 148

DB 1117 AATAAAAAAGGAAACAAACAAATAAGAAAAATGAAAAAGAAAAATAAAAAAAG 1176

QY 149 SerAsnAlaAsnAlaHisPheThrSerLysArgLysAspThrLysArgArgGlnGlu 168

DB 1177 -----GAAAAAGAAAAATAAAAAAGAAAAAGTAAAGAAAAAGAAAGAA 1209

QY 169 ArgHisIleLysGluGluGlnAspLysGluMetThrSerAlaLysGln 184

DB 1210 AAAAATAAAAAAGGAAAGAAAAAGTAAAGAAAAAGGAAAAAGAA 1257

RESULT 31

AAAD24393

ID AAD24393 standard; cDNA; 2220 BP.

XX AAD24393;

AC AAD24393;

XX 07-MAR-2002 (first entry)

DT Human RNA metabolism protein-20 (RMEP-20) cDNA.

DE Human; RNA metabolism protein-20; RMEP-20; gout; nervous system disorder;

XX autoimmunity; inflammatory; cell proliferative; developmental; thyroiditis;

CC gene therapy; epilepsy; dementia; stroke; Alzheimer's disease; amnesia;

CC Parkinson's disease; prion disease; insomnia; endocrine disorder; AIDS;

CC Acquired Immune Deficiency Syndrome; mental disorder; allergy; anaemia;

CC asthma; atherosclerosis; Crohn's disease; rheumatoid arthritis; vaccine;

KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis; drug; side effect; cancer; central nervous system; cardiovascular; KW gastrointestinal; respiratory system; single nucleotide polymorphism; KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
XX WO200218632-A2.
XX
XX
XX PD 07-MAR-2002.
XX
XX
XX PF 01-SEP-2001; 2001WO-EP10074.
XX
XX PR 01-SEP-2000; 2000DE-1043826.
XX
XX PR 05-SEP-2000; 2000DE-1044543.
XX
XX PA (EPiG-) EPIGENOMICS AG.
XX
XX PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX WPI; 2002-371829/40.
XX
XX PT Determining the degree of cytosine methylation in genomic DNA, useful for diagnosis and prognosis, comprises selective hybridization of PT amplicons from chemically treated DNA.
XX
XX PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
XX CC This invention describes a novel method for determining the degree of CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a CC genomic sample of DNA. The sample is treated chemically to convert CC cytosine (C) but not methylated C, to uracil, then part of the genomic CC DNA that contains the target C is amplified to form a labeled amplicon. CC The amplicon is hybridised to two classes, each with at least one CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers CC and the degree of hybridisation to both classes is determined from the CC label on the amplicon. From the ratio of labels hybridised to the two CC classes of oligomers, the degree of methylation is calculated. The method CC is used: (i) for diagnosis and/or prognosis of side effects of CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders CC of the central nervous, cardiovascular, gastrointestinal and respiratory CC systems etc., particularly by detecting mutations or single nucleotide CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue CC types and for investigating cell differentiation. The method allows the CC methylation status of many C residues to be determined simultaneously. CC AB013410-AB054121 represent genomic DNA sequences used to illustrate the CC method for determining the degree of cytosine methylation described in CC the disclosure of the invention.
XX
XX SQ Sequence 886 BP; 114 A; 123 C; 309 G; 340 T; 0 other;

Alignment Scores:
Pred. No.: 0.0565 Length: 886
Score: 101.00 Matches: 24
Percent Similarity: 52.70% Conservative: 15
Best Local Similarity: 32.43% Mismatches: 33
Query Match: 10.01% Indels: 2
DB: 24 Gaps: 1

US-09-868-352-23 (1-192) x ABQ35134 (1-886)

Qy 112 AsnProAsnProArgAspIleLysPheAspGluArgProHisLysArgArgLysSer 131
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 439 AACAGAAATCCAGCTACTGACGTATCGTATCTCTGCTCCCAACGTAACGACTAAATCA 380
Qy 132 ArgSerLysSerGlnSerSerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAla 151
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 379 ACCGCTCTCCAAATACCGCGGACACACACGCGTCTCGCTAAACCC 320
Qy 152 AsnAlaHisPheThrSerLysLysArgLysAspThrLysArgArgGlnGluArgHisIle 171
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 319 AATAATCGACCTCTTAACCAAAAGCAACGCACTCGTCCGAACACGTACAAAT--- 263

Qy 172 LysGluGluGlnAspLysGluMetThrSerAlaLysGlnHis 185
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 262 --AACCAATACGATAAAATACACAACTCAAAAAAACAAC 224
RESULT 29
ABQ35135
ID ABQ35135 standard; DNA; 886 BP.
XX
XX AC ABQ35135;
XX
XX DT 12-JUL-2002 (first entry)
XX
XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 21726.
XX
XX KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis; drug; side effect; cancer; central nervous system; cardiovascular; KW gastrointestinal; respiratory system; single nucleotide polymorphism; KW SNP; cell differentiation; ds.
XX
XX OS Homo sapiens.
XX
XX PN WO200218632-A2.
XX
XX PD 07-MAR-2002.
XX
XX PF 01-SEP-2001; 2001WO-EP10074.
XX
XX PR 01-SEP-2000; 2000DE-1043826.
XX
XX PR 05-SEP-2000; 2000DE-1044543.
XX
XX PA (EPiG-) EPIGENOMICS AG.
XX
XX PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX WPI; 2002-371829/40.
XX
XX PT Determining the degree of cytosine methylation in genomic DNA, useful for diagnosis and prognosis, comprises selective hybridization of PT amplicons from chemically treated DNA.
XX
XX PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
XX CC This invention describes a novel method for determining the degree of CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a CC genomic sample of DNA. The sample is treated chemically to convert CC cytosine (C) but not methylated C, to uracil, then part of the genomic CC DNA that contains the target C is amplified to form a labeled amplicon. CC The amplicon is hybridised to two classes, each with at least one CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers CC and the degree of hybridisation to both classes is determined from the CC label on the amplicon. From the ratio of labels hybridised to the two CC classes of oligomers, the degree of methylation is calculated. The method CC is used: (i) for diagnosis and/or prognosis of side effects of CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders CC of the central nervous, cardiovascular, gastrointestinal and respiratory CC systems etc., particularly by detecting mutations or single nucleotide CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue CC types and for investigating cell differentiation. The method allows the CC methylation status of many C residues to be determined simultaneously. CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the CC method for determining the degree of cytosine methylation described in CC the disclosure of the invention.
XX
XX SQ Sequence 886 BP; 340 A; 309 C; 123 G; 114 T; 0 other;

Alignment Scores:
Pred. No.: 0.0565 Length: 886
Score: 101.00 Matches: 24
Percent Similarity: 52.70% Conservative: 15
Best Local Similarity: 32.43% Mismatches: 33
Query Match: 10.01% Indels: 2
DB: 24 Gaps: 1

KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
 KW cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
 KW antifungal; vulnerary; antiulcer; ss.
 XX Homo sapiens.
 XX WO200157188-A2.
 PN 09-AUG-2001.
 PD
 PF 05-FEB-2001; 2001WO-US03800.
 XX 03-FEB-2000; 2000US-0496914.
 PR 27-APR-2000; 2000US-0560875.
 XX (HYSE-) HYSEQ INC.
 XX PA
 XX PI Tang YT, Liu C, Drmanac RT;
 XX WI: 2001-457740/49.
 DR P-PSDB; ABB11354.
 XX Human proteins and DNA encoding sequences useful for preventing,
 PT treating or ameliorating a medical condition in a mammalian subject
 PT e.g. arthritis and cancer -
 XX
 PS Claim 1; Page 469-470; 1963pp; English.
 XX Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
 CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
 CC invention also relates to vectors and recombinant host cells comprising a
 CC nucleotide of the invention, methods of producing the novel polypeptides,
 CC antibodies against the polypeptides, methods of detecting the nucleotides
 CC or polypeptides in a sample, and methods of identifying compounds which
 CC bind to polypeptides of the invention. Although novel, many of the
 CC polypeptides of the invention have homology to known proteins, thereby
 CC giving an insight into their probable biological activities, and hence
 CC potential therapeutic applications. The polypeptides of the invention may
 CC have various activities, including cytokine, cell proliferation or cell
 CC differentiation activities; stem cell growth factor activity;
 CC haematopoiesis regulatory activity; tissue growth activity;
 CC immunomodulatory activity; activin- or inhibin-related activities;
 CC chemotactic or chemokinetic activities; haemostatic, thrombotic or
 CC thrombolytic activities; receptor or ligand activities; or may be
 CC involved in oncogenesis, cancer cell proliferation or metastasis.
 CC Depending on their biological activities, polypeptides and nucleotides of
 CC the invention are useful for preventing, treating or ameliorating medical
 CC conditions, e.g., by protein or gene therapy. Such conditions include
 CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell
 CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
 CC proliferative retinopathy, atherosclerosis, coronary heart disease,
 CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
 CC vascular growth. Polypeptides involved with tissue regeneration and
 CC repair (or nucleic acids encoding them) may be used to promote wound
 CC healing (e.g., of burns, incisions and ulcers), while those with
 CC immunomodulatory activities may be used in the treatment of viral,
 CC bacterial and fungal infections in addition to immune disorders.
 CC Polypeptides with growth factor activity may be used in cell cultures to
 CC promote cell growth. For example, such polypeptides may be used to
 CC manipulate stem cells in culture to give rise to neuroepithelial cells
 CC that can be used to augment or replace cells damaged by illness,
 CC autoimmune disease or accidental damage. The polypeptides and nucleotides
 CC may also be used in the diagnosis of the above conditions, and in drug
 CC screening techniques. The present sequence represents a cDNA encoding a
 CC novel human polypeptide of the invention.
 XX
 SQ Sequence 1909 BP; 687 A; 340 C; 461 G; 415 T; 6 other;

Alignment Scores:

Pred. No.: 0.105 Length: 1909
 Score: 102.50 Matches: 33
 Percent Similarity: 48.96% Conservative: 14
 Best Local Similarity: 34.38% Mismatches: 36

Query Match: 10.16% Indels: 14
 DB: 22 Gaps: 4
 US-09-868-352-23 (1-192) x ABA08598 (1-1909)
 QY 103 PheGlyCysAla-----TyrPheValLeuGluAsnProAsnPro 115
 DB 933 TTTAATTGTTCTGAGCTTACACTGCAAAAGTGATTTTCTCCAGAGTCTCGAAGAGC 992
 QY 116 ArgAspIleLysPheAspGluArgProHisLysArgGlnLysSerArgSerLysSer 135
 DB 993 AATGAAGAAAAGCGGTGCGATCTGTTCCCATATCTCGTCAAAATCCAGGTCT----- 1046
 QY 136 GlnSerSerLysSerGlnThrArgAsnAsnArgSerGlnSer-----AsnAlaAsnAla 153
 DB 1047 ---AGCTCAAAATCCCATTTCTAGAAGGAAAGACATCACAAATCAACACAGGAGTAGATCC 1103
 QY 154 HisPheThrSerLysLysArg---LysAspThrLysArgGlnGlnArgHisLysLys 172
 DB 1104 CATAATAGATCACGTTCAAGACAGAAAGACAGACGTAGATCTTAAGAGCCCAATAAAAA 1163
 QY 173 GluGluGlnAspLysGluMetThrSerAlaLysGlnHisLeuLeuPhe 188
 DB 1164 CGCTCTAAATCAAGGAGAG-ACGGAAGTCAAGGAGTCTGTCGCATTC 1210
 RESULT 25
 AAD39557
 ID AAD39557 standard; DNA; 2695 BP.
 XX AC AAD39557;
 XX DT 04-OCT-2002 (first entry)
 XX DE Human SR-cyp DNA.
 XX KW Human; antisense: SR-cyp; Clk-associated RS cyclophilin; inflammation;
 KW hyperproliferative disorder; cancer; prophylaxis; infection; therapy;
 KW tumour; CARs-cyp; gene; ds.
 XX OS Homo sapiens.
 XX FH Key Location/Qualifiers
 FT CDS 158..2422
 FT /*tag= a
 FT /product= "Human SR-cyp protein"
 XX PN WO200236809-A2.
 XX PD 10-MAY-2002.
 XX PF 30-OCT-2001; 2001WO-US47335.
 XX PR 03-NOV-2000; 2000US-0706197.
 XX XX (ISIS-) ISIS PHARM INC.
 XX PA (COLD-) COLD SPRING HARBOR LAB.
 XX PI Bennett CF, Spector DL, Wyatt JR;
 XX WPI: 2002-479763/51.
 DR P-PSDB; AAE24596.
 XX Novel antisense compounds targeted to nucleic acids encoding SR-cyp,
 PT Clk-associated RS cyclophilin for modulating the gene expression and
 PT treating hyperproliferative disorders such as cancer -
 XX
 PS Example 13; Page 95-99; 117pp; English.
 XX
 CC The invention relates to antisense compounds targeted to a nucleic
 CC acid molecule encoding human SR-cyp (Clk-associated RS cyclophilin)
 CC to inhibit its expression. SR-cyp is also referred to as CARs-cyp.
 CC Antisense compounds of the invention are used for treating diseases
 CC or conditions associated with SR-cyp. The diseases treated include


```
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI: 2001-442253/47.
DR P-PSDB; AAM39441.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Claim 1; SEQ ID NO 800; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
XX Sequence 1826 BP; 694 A; 338 C; 453 G; 341 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 0.0206 Length: 1826
Score: 108.50 Matches: 32
Percent Similarity: 47.92% Conservative: 14
Best Local Similarity: 33.33% Mismatches: 37
Query Match: 10.75% Indels: 13
DB: 22 Gaps: 4
US-09-868-352-23 (1-192) x AAI58597 (1-1826)
QY 103 PheGlyCysAla-----TyrPheValLeuGluAsnProAsnPro 115
Db 694 TTTAATGTCTCAGCTTACACTGCAAGTGATTTTCTCCAGAGTCTGGAAGAGC 753
QY 116 ArgAspIleLysPheAspGluArgProHisLysArgArgLysSerArgSerLysSer 135
Db 754 AATGAAGAAAGACGGCTGCTCGTCCCTCACTCGCTCAAAATCCAGGTCT----- 807
QY 136 GlnSerSerLysSerGlnThrArgAsnAsnArgSerGlnSer-----AsnAlaAsnAla 153
Db 808 ---AGTCAAAATCCCATTTCTAGAAGGAAAGAGATCAATCAAAACACAGAGTAGATCC 864
QY 154 HisPheThrSerLysLysArg---LysAspThrLysArgArgGlnGluArgHisIleLys 172
Db 865 CATAATAGATCAGTTCAACACAGAGAAGACAGAGTAGATCTTAGAGCCCAACATAAAAA 924
QY 173 GluGluGlnAspLysGluMetThrSerAlaLysHisLeuLeuPhe 188
Db 925 CGCTCTAAATCAAGGGGAGAGACGGAAGTCAAGAGTCGTTCGCATTC 972
RESULT 22
ID AAS79353
XX
XX AAS79353 standard; cDNA; 1808 BP.
XX
XX AAS79353;
XX
```

```
DT 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #15157.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
DR P-PSDB; ABG15166.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 1; SEQ ID NO 15157; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1808 BP; 541 A; 437 C; 549 G; 281 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 0.0508 Length: 1808
Score: 105.00 Matches: 36
Percent Similarity: 42.94% Conservative: 34
Best Local Similarity: 22.09% Mismatches: 65
Query Match: 20.41% Indels: 28
DB: 23 Gaps: 5
US-09-868-352-23 (1-192) x AAS79353 (1-1808)
QY 46 ValThrValPheGlyGlnArgPheSerGluIleLeuLeuLysTyrAsp----- 61
Db 652 CTGACACAGTTGAAGGAGTCTTTTCAACAAGTCCAAATAGAAAGAGATGAGTATTCGAA 711
QY 62 PheIleValGlyAspTrpGlyGlyAsnGluGlnLeuArgGlyPheTyrLysAspAla 81
Db 712 CATCTAAAGGAGAGAGGGCCCGTGGCAGCAGGAGTGAAGAAATGTCGAGGAGATT 771
```

```
Qy 54 SerGluIleLeuLeuLysTyrAspPheIleValGlyAspTTPGlyAsnGluInLeuArg 73
Db 86 TCCGACATTTTAAATACGACTACATCTCGGGGATTGGGATACAGCCAGCTGAGG 145
Qy 74 LeuArgGlyPheTyrLysAspAlaSer---ThrIleArgLysAsnSerArgIleSerArg 92
Db 146 CTGAAGAGCTTTTGTGACGATCAAAATCAAAAGGCCACATTTGTACGAAGATCAGCAG 205
Qy 93 LeuGluAspTyrIleLysTyrCysAsnPheGlyCysAlaTyrPheValLeuGlu 111
Db 206 CTGTATGATATATATTACGACTACTGCAATTTCCGATGTCCTACTTTGTTTAAAA 262
RESULT 20
AAX84569
ID AAX84569 standard; cdna to mRNA; 6145 BP.
XX
AC AAX84569;
XX
DT 13-SEP-1999 (first entry)
XX
DE Biorhythm marker gene.
XX
KW Biorhythm marker gene; circadian rhythm; ss.
XX
OS Rattus sp.
XX
PN JP11169185-A.
XX
PD 29-JUN-1999.
XX
PF 12-DEC-1997; 97JP-0362890.
XX
PR 12-DEC-1997; 97JP-0362890.
XX
PA (AGEN ) AGENCY OF IND SCI & TECHNOLOGY.
PA (SAKA ) OTSUKA PHARM CO LTD.
XX
DR WPI; 1999-422623/36.
DR P-PSDB; AAY22206.
XX
PT Structure and function of a new biorhythm marker gene - and protein
PT encoded by it
XX
PS Claim 7; Page 14-16; 23pp; Japanese.
XX
CC This sequence encodes the biorhythm marker protein of the invention. The
CC protein is also referred to as a circadian rhythm protein.
XX
SQ Sequence 6145 BP; 1849 A; 1110 C; 1364 G; 1822 T; 0 other;
Alignment Scores:
Pred. No.: 0.0277 Length: 6145
Score: 113.50 Matches: 55
Percent Similarity: 37.16% Conservative: 42
Best Local Similarity: 21.07% Mismatches: 73
Query Match: 11.25% Indels: 91
DB: Gaps: 11
US-09-868-352-23 (1-192) x AAX84569 (1-6145)
Qy 8 GluMetLeuAsnTyrAsnLysTyrProGlyProGlnPheIleHisPheGluAsnIleVal 27
Db 2788 AAATGTTAAATGTAAGCAGGTTCCTTCCTCAAGTTTAAATTTTCAGACAGAGAA 2847
Qy 28 LysSerAspIleGlu----- 33
Db 2848 AGGCACTCTAATTTAGAAAACCAAGATATTTCAGTTTCTAGACTAAAGGAAAACTTACT 2907
Qy 34 -----PheGlnLeuValIleAsnGluLysSerAlaPhe-----AspValThr 47
Db 2908 CAAATTTTGTGTGCTAAATACATTTTAAACTCCACTCTGAACCTCTCTATGTGATTATCT 2967
Qy 48 ValPheGlyGlnArgPheSerGluIleLeuLysTyrAspPhe----- 62
```

```
Db 2968 TTCATT-----AGATTTCCTAAATTAATACTTGAAATTTAGTTTCGGTCTTAACAAAATA 3021
Qy 63 -----IleValGlyAsp 66
Db 3022 AATCAGGAAAGTCTTACGCTCCTACTGCTCTAGGAATTGCAGACATGTTTAACTCTGCGT 3081
Qy 67 TrpGlyAsnGluInLeuArgGlyPheTyrLysAspAlaSerThrIleArgLys 86
Db 3082 TATGCGACAGACAGAAAT---CTAGACAGCTGTTTATAATATCATCTCGTTTATATGTAC 3138
Qy 87 AsnSerArgIleSerArgLeuGluAsp-----TyrIleLys 98
Db 3139 TGCTGCAGGGTGGCTGCAGCTCAACGAGTTTATGCAATGACCTTCTTGGGATGTTCTGAAG 3198
Qy 99 GluTyrCysAsnPheGly----- 104
Db 3199 GATAATTGCACAGGAGGAGGATGTACAGAGAGTAGGCCCTTGCACTATATGTGTATAT 3258
Qy 105 -----CysAlaTyr----- 107
Db 3259 TCCACTTGTGCCTATTATTAACTGGGATCTTTAATTGTTCTGAGCTTACGCTGCAAGT 3318
Qy 108 -----PheValLeuGluAsnProArgAspIleLysPheAspGluArgPro 125
Db 3319 GGTGTTTCTCCAGAGTCTGGAAAGAGCAATCAAGGAAGGCGTCTCGTTCC 3378
Qy 126 HisLysArgArgLysSerArgSerLysSerGlnSerLysSerGlnThrArgAsnAsn 145
Db 3379 CACACTCGCTCAAAGTCCAGTCT-----AGCTCAAAGTCCCATTTCTAGAAGGAAG 3429
Qy 146 ArgSerGlnSer-----AsnAlaAsnAlaHisPheThrSerLysLysArg---LysAsp 162
Db 3430 AGATCGCAGTCAAAAGCAGGAGTAGATCCCATACAGGTCCAGTCAAGACAGAAAGAT 3489
Qy 163 ThrLysArgArgGlnGluArgHisIleLysGluGlnAspLysGluMetThrSerAla 182
Db 3490 AGAGCTAGATCCAAGAGCCACATAAGAACGCTTAAGTCTAGGAGAGCGCGAAGTCA 3549
Qy 183 Lys 183
Db 3550 AGG 3552
RESULT 21
AAX158597
ID AAX158597 standard; cdna; 1826 BP.
XX
AC AAX158597;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 800.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW ankyrotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
```

PI Berka R, Clausen IG;
XX WPI; 2002-416684/44.
XX Monitoring differential expression of several genes in first Bacillus
PT cell relative to expression of same genes in one or more second
PT Bacillus cells, by using substrate containing Bacillus genomic
PT sequenced tag array -
XX
XX Claim 11; SEQ ID NO 7936; 200pp; English.
XX
XX The invention describes a method of monitoring differential expression of
CC genes in a first Bacillus cell relative to expression of the genes in
CC other Bacillus cells, comprising hybridising labelled nucleic acid probes
CC isolated from Bacillus cells to a substrate containing array of Bacillus
CC genomic sequenced tags (GST), examining the array, and determining
CC relative gene expression by an observed hybridisation reporter signal of
CC a spot in the array. The method is useful for measuring the expression of
CC genes in a first Bacillus cell relative to expression of the same genes
CC in one or more second Bacillus cells. The method is useful for monitoring
CC global expression of several genes from a Bacillus cell, discovering new
CC genes, identifying possible functions of unknown open reading frames and
CC monitoring gene copy number variation and stability. Monitoring changes
CC in expression of genes may be used to provide a representation of the way
CC in which Bacillus cells adapt to changes in culture conditions,
CC environmental stress or other physiological provocation. Extensive
CC follow-up characterisation is unnecessary, when one spot on an array
CC equals one gene or one open reading frame, since sequence information is
CC available. This sequence represents a genomic sequence tag (GST) used in
CC the method of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 285 BP; 93 A; 51 C; 66 G; 74 T; 1 other;
XX
Alignment Scores:
Pred. No.: 1.46e-17 Length: 285
Score: 232.50 Matches: 46
Percent Similarity: 69.32% Conservative: 15
Best Local Similarity: 52.27% Mismatches: 22
Query Match: 23.04% Indels: 5
DB: 24 Gaps: 2
US-09-868-352-23 (1-192) x ABK80645 (1-285)
QY 26 IleValLysSerAspApilLeuGluPheGlnLeuValIleAsnGluLysSerAlaPheAsp 45
Db 1 ATGGTGACGATTCAGCGATTCATGATATGAGTCTCGAAACGAGCGCATGGTGGAAAT 60
QY 46 ValThrValPheGlyGlnArgPheSerGluIleLeuLeuLysTyrAspPheIleValGly 65
Db 61 GAGGAGCGATTCAAAGCGCGGTACAGCGATGATCAACAAATATGATTATCATCGTCGCG 120
QY 66 AspTrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAsp-----AlaSer 82
Db 121 GATTGGGGCTACACCAACGCGGTGCGGGCTTTTGTGATGACGACAAACAAAAATCA 180
QY 83 ThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyrCysAsn 102
Db 181 ACATTCGATAAA-----AAATAGACACGTTACCCGACTATCTTTACGATACTGCAAT 234
QY 103 PheGlyCysAlaTyrPheValLeu I10
Db 235 TTTGGTTCGCTTATTGTCCTG 258
RESULT 19
ABK75854
ID ABK75854 standard; DNA; 273 BP.
XX
AC ABK75854;
XX

DT 13-AUG-2002 (first entry)
XX
DE Bacillus licheniformis genomic sequence tag (GST) #3145.
XX
KW Differential gene expression; genomic sequenced tag; GST;
KW altered culture condition; environmental stress;
KW physiological provocation; ds.
XX
OS Bacillus licheniformis.
XX
PN WO200229113-A2.
XX
PD 11-APR-2002.
XX
PF 05-OCT-2001; 2001WO-US31437.
XX
PP 06-OCT-2000; 2000US-0680598.
PR 27-MAR-2001; 2001US-279526P.
XX
XX (NOVO) NOVOZYMES BIOPECH INC.
XX (NOVO) NOVOZYMES AS.
XX Berka R, Clausen IG;
XX WPI; 2002-416684/44.
XX
XX Monitoring differential expression of several genes in first Bacillus
PT cell relative to expression of same genes in one or more second
PT Bacillus cells, by using substrate containing Bacillus genomic
PT sequenced tag array -
XX
XX Claim 4; SEQ ID NO 3145; 200pp; English.
XX
XX The invention describes a method of monitoring differential expression of
CC genes in a first Bacillus cell relative to expression of the genes in
CC other Bacillus cells, comprising hybridising labelled nucleic acid probes
CC isolated from Bacillus cells to a substrate containing array of Bacillus
CC genomic sequenced tags (GST), examining the array, and determining
CC relative gene expression by an observed hybridisation reporter signal of
CC a spot in the array. The method is useful for measuring the expression of
CC genes in a first Bacillus cell relative to expression of the same genes
CC in one or more second Bacillus cells. The method is useful for monitoring
CC global expression of several genes from a Bacillus cell, discovering new
CC genes, identifying possible functions of unknown open reading frames and
CC monitoring gene copy number variation and stability. Monitoring changes
CC in expression of genes may be used to provide a representation of the way
CC in which Bacillus cells adapt to changes in culture conditions,
CC environmental stress or other physiological provocation. Extensive
CC follow-up characterisation is unnecessary, when one spot on an array
CC equals one gene or one open reading frame, since sequence information is
CC available. This sequence represents a genomic sequence tag (GST) used in
CC the method of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 273 BP; 89 A; 48 C; 61 G; 75 T; 0 other;
XX
Alignment Scores:
Pred. No.: 1.89e-16 Length: 273
Score: 222.50 Matches: 42
Percent Similarity: 74.68% Conservative: 17
Best Local Similarity: 53.16% Mismatches: 19
Query Match: 22.05% Indels: 1
DB: 24 Gaps: 1
US-09-868-352-23 (1-192) x ABK75854 (1-273)
QY 34 PheGlnLeuValIleAsnGluLysSerAlaPheAspValThrValPheGlyGlnArgPhe 53
Db 26 TTTGAAGTCGTCAGAGATGTGAAAAACGGTTTTAACCAAGACGGTTTAAAGCCAGATAC 85

CC anti-Listeria vaccines.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1234 BP; 402 A; 195 C; 288 G; 349 T; 0 other;
 Alignment Scores:
 Pred. No.: 1-62e-17 Length: 1234
 Score: 239.50 Matches: 43
 Percent Similarity: 77.38% Conservative: 22
 Best Local Similarity: 51.19% Mismatches: 14
 Query Match: 23.74% Indels: 5
 DB: 24 Gaps: 2
 US-09-868-352-23 (1-192) x ABQ68718 (1-1234)
 Qy 31 AspIleGluPheGlnLeuValIleAsnGluLysSerAlaPheAspValThrValPheGly 50
 Db 541 GATTTAAACTACGAGATTATCACCAAATTCGCGACGCTTTTGACGAGGAGAAGTTAAAT 600
 Qy 51 GlnArgPheSerGluIleLeuLeuLysTyrAspPheIleValGlyAspTrpGlyAsnGlu 70
 Db 601 GAACGGTTTAGCGATATCTTGGCGGATATGATTATATAGTAGCGGACTGGGGTTACGAT 660
 Qy 71 GlnLeuArgLeuArgGlyPheTyrLysAspAlaSerThrIleArgLys-----Asn 87
 Db 661 CAACTCGCTCTCAAGGCTTTTGAAGACGATAAC-----CGAAAGCCGCATACGAC 714
 Qy 88 SerArgIleSerArgLeuGluAspTyrIleLysGluTyrCysAsnPheGlyCysAlaTyr 107
 Db 715 AACAAATAGCACGTAAAGAATATATTACGAATACTGCAATTTTGGTTGTGCGTAT 774
 Qy 108 PheValLeuGlu 111
 Db 775 TTTGTTATTAAA 786
 RESULT 17
 ABQ70820
 ID ABQ70820 standard; DNA; 2690 BP.
 XX
 AC ABQ70820;
 XX
 DT 29-AUG-2002 (first entry);
 XX
 DE Listeria monocytogenes 4b contig DNA sequence #762.
 XX
 KW Antibacterial; Listeria; food contamination; mutational analysis;
 KW infection; ds.
 XX
 OS Listeria monocytogenes 4b.
 XX
 PN WO200228891-A2.
 XX
 PD 11-APR-2002.
 XX
 PE 04-OCT-2001; 2001WO-FR03061.
 XX
 PR 04-OCT-2000; 2000FR-0012697.
 XX
 PA (INSP) INST PASTEUR.
 PA (CNRS) CNRS CENT NAT RECH SCI.
 XX
 PI Kunst F, Glaser P;
 XX
 DR WPI; 2002-332479/37.
 XX
 PT New genomic sequences from Listeria species, useful for detection,
 PT treatment and prevention of infection, also related polypeptides,
 PT antibodies and modulators
 XX
 PS Claim 14; SEQ ID 3633; 180pp; French.
 XX

CC The present invention relates to nucleic acid sequences
 CC (ABQ67188-ABQ71212) from Listeria sp. The sequences are useful as probes
 CC and primers for identification and/or detection of Listeria (e.g. as
 CC contaminants in foods, or mutational analysis) and for analysis of
 CC gene expression. Proteins encoded by the nucleic acid sequences can be
 CC used to screen for compounds that modulate gene expression, replication
 CC and pathogenicity of Listeria (potential therapeutic agents), also for
 CC treating infections by Listeria, and are useful as immunogens in
 CC anti-Listeria vaccines.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 2690 BP; 852 A; 456 C; 629 G; 753 T; 0 other;
 Alignment Scores:
 Pred. No.: 4.54e-17 Length: 2690
 Score: 239.50 Matches: 43
 Percent Similarity: 77.38% Conservative: 22
 Best Local Similarity: 51.19% Mismatches: 14
 Query Match: 23.74% Indels: 5
 DB: 24 Gaps: 2
 US-09-868-352-23 (1-192) x ABQ70820 (1-2690)
 Qy 31 AspIleGluPheGlnLeuValIleAsnGluLysSerAlaPheAspValThrValPheGly 50
 Db 541 GATTTAAACTACGAGATTATCACCAAATTCGCGACGCTTTTGACGAGGAGAAGTTAAAT 600
 Qy 51 GlnArgPheSerGluIleLeuLeuLysTyrAspPheIleValGlyAspTrpGlyAsnGlu 70
 Db 601 GAACGGTTTAGCGATATCTTGGCGGATATGATTATATAGTAGCGGACTGGGGTTACGAT 660
 Qy 71 GlnLeuArgLeuArgGlyPheTyrLysAspAlaSerThrIleArgLys-----Asn 87
 Db 661 CAACTCGCTCTCAAGGCTTTTGAAGACGATAAC-----CGAAAGCCGCATACGAC 714
 Qy 88 SerArgIleSerArgLeuGluAspTyrIleLysGluTyrCysAsnPheGlyCysAlaTyr 107
 Db 715 AACAAATAGCACGTAAAGAATATATTACGAATACTGCAATTTTGGTTGTGCGTAT 774
 Qy 108 PheValLeuGlu 111
 Db 775 TTTGTTATTAAA 786
 RESULT 18
 ABK80645
 ID ABK80645 standard; DNA; 285 BP.
 XX
 AC ABK80645;
 XX
 DT 13-AUG-2002 (first entry)
 XX
 DE Bacillus clausii genomic sequence tag (GST) #3488.
 XX
 KW Differential gene expression; genomic sequenced tag; GST;
 KW altered culture condition; environmental stress;
 KW physiological provocation; ds.
 XX
 OS Bacillus clausii.
 XX
 PN WO200229113-A2.
 XX
 PD 11-APR-2002.
 XX
 PF 05-OCT-2001; 2001WO-US31437.
 XX
 PR 06-OCT-2000; 2000US-0680598.
 PR 27-MAR-2001; 2001US-279526P.
 XX
 PA (NOVO) NOVOZYMES BIOTECH INC.
 PA (NOVO) NOVOZYMES AS.
 XX

OS	Staphylococcus epidermidis.
XX	
XX	WO200134809-A2.
XX	
XX	17-MAY-2001.
XX	
XX	09-NOV-2000; 2000WO-US30782.
PF	
XX	
PR	09-NOV-1999; 99US-0164258.
XX	
PA	(GLAX) GLAXO GROUP LTD.
XX	
XX	Kimmerly WJ;
PI	
XX	
XX	WPI: 2001-316495/33.
DR	
DR	P-PSDB; AAG82114.
XX	
PT	Nucleic acids encoding polypeptides from staphylococcus epidermidis,
PT	useful for vaccinating against infections, e.g. endocarditis -
XX	
PS	Claim 8; Page 376; 2188pp; English.
XX	

XX WPI: 2001-316495/33.
 DR p-PSDB: AAG82114.
 XX
 XX Nucleic acids encoding polypeptides from staphylococcus epidermidis,
 PT useful for vaccinating against infections, e.g. endocarditis -
 XX
 XX Claim 8: Page 376: 2188pp: English.
 PS

XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides

CC (II), given in AAG81454 to AAG83120, from *Staphylococcus epidermidis*.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids(I) may be used to produce the

CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce hosts cells which express the
CC S. epidermidis polypeptides (II) (and/or nucleic acids) may then be
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be

used to vaccinate subjects and to raise antibodies against the bacteria. The polypeptides may also be used to assay for other inhibitors of their

CC activity and therefore identify compounds that may be used for the
CC treatment of *S. epidermidis* infections, e.g. endocarditis. AAH53971 to
CC AAH5090 represent specifically claimed *S. epidermidis* genomic DNA

CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.

CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,

CC however the sequence listing only goes up to SEQ ID NO:4434 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.

AA
SQ Sequence 390 BP; 158 A; 41 C; 67 G; 124 T; 0 other;

Alignment scores:

Pred. No.:	1.25e-18	Length:	390
Score:	243.50	Matches:	45

Conservative:	24
Percent Similarity:	66.32%
Best Local Similarity:	44.55%
Mismatches:	29
Indels:	3
Query Match:	24.13%

DB:	22	Gaps:	2
US-09-868-352-23 (1-192) x AAH52964 (1-390)			

QY 26 IlevallysSerAspaspIleGluPheGlnLeuValIleAsnGluLysSerAlaPheAsp 45
 ::::||||| |||||:::|||||::: :::

DB 1 ATGATTAAAGTCGACCAACAATATTTTGAATTGTATAGAGAAATATAGAGAAATGTTTGTAT 60

Ov 46 ValThrValPheGlyGlnArgPheSerGluIleLeuLeuIysTyrAspPheIleValGly 65

Db 61 GAGGAAATATTTTCAGCTAGGTATTCGGATATATTAGACAAATATGATTATGTCGTAGGT 120

DQ 66 AsptPrpGlyAsnGlucInLeuArgLeuArgGlyPheTyrLysASPAlaSerThrIleArg 85
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DP 121 GACTATGGTTACGATCAAAATTACGGTTAAAAGCATTTTTATAAAGATACGTAAATAAAAAAGCCA 180

QY 86 Lys---AsnSerArgLeuGluAspTyrIleLysGluTyrCysAsnPheGly 104
 ::: ||| ::: |||| |||| ||||
QV

DB 181 GAAATAAGCTAAACGATTTCCTCAAGTATACACAGATATATACACAGATATATGTAATTTTGGT 240

Qv 105 CvsAlatvPheValIeuGluAsnProAsnProArgAspIle-----LvspheAspAsp 122

[illegible]

KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.
XX Staphylococcus aureus.
XX
FH Key Location/Qualifiers
FT misc_feature 1141..1200
FT /*tag= a
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 2941..3000
FT /*tag= b
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 4741..4800
FT /*tag= c
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 6541..6600
FT /*tag= d
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
XX
PN EF786519-A2.
XX
PD 30-JUL-1997.
XX
XX
PF 07-JAN-1997; 97EP-0100117.
XX
XX 05-JAN-1996; 96US-0009861.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;
XX Rosen CA;
XX
XX WPI; 1997-374922/35.
XX
XX Polynucleotide(s) and proteins derived from Staphylococcus aureus -
XX stored on computer readable medium and used in the production of
XX anti-S.aureus vaccines
XX
XX Claim 1; Page 958-962; 3271pp; English.
XX
XX This sequence represents one of 5191 Staphylococcus aureus DNA sequences
XX of the invention. The DNA sequences are recorded on a computer readable
XX medium, preferably selected from a floppy or hard disk, random access
XX memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
XX the S.aureus DNA sequences allows putative functions to be assigned so
XX that protein-encoding or regulatory regions of commercial, therapeutic or
XX industrial importance can be obtained. Specifically, sequences which are
XX likely to encode antigens have been identified and these polypeptides can
XX be used in a vaccine composition against S.aureus infection. The
XX polypeptides can also be used in a kit for the immunodetection of
XX S.aureus in a sample. S.aureus is implicated in numerous human diseases,
XX including cellulitis, eyelid infections, food poisoning, osteomyelitis,
XX skin and surgical wound infections, scalded skin syndrome, toxic shock
XX syndrome, etc. Organisms transformed with the DNA sequences can be used
XX for recombinant production of the polypeptides. The new DNA sequences
XX (and their fragments) are useful as primers or probes for isolating
XX homologues of any of the S.aureus DNA sequences contained on the
XX computer readable medium.

SQ Sequence 7769 BP; 2697 A; 1099 C; 1425 G; 2307 T; 241 other;
Alignment Scores:
Pred. No.: 1.35e-17 Length: 7769
Score: 249.50 Matches: 50
Percent Similarity: 62.28% Conservative: 21
Best Local Similarity: 43.86% Mismatches: 30
Query Match: 24.73% Indels: 13
DB: 18 Gaps: 2
US-09-868-352-23 (1-192) x AAV74525 (1-7769)
QY 24 GluAsnIleValLysSerAspIleGluPheGlnLeuValIleAsnGluLysSerAla 43
Db 6606 AGAATTTGATAAAAGTAGATCAACATTACTTTGAATTAATAAGAAATTCGCGAATGT 6665
QY 44 PheAspValThrValPheGlyGlnArgPheSerGluIleLeuLeuLysTytraspPheIle 63
Db 6666 TTTAATGAAGACAATTTTGTCTAGGTATTTCAGATATTTTAGATAAATATGATTACATA 6725
QY 64 ValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPheTytrLysAspAlaSer--- 82
Db 6726 GTTGTGCTACTGTTACGATCAATTTACGATTAAGAGTTTTCACAAAGATTCTTAATAAA 6785
QY 83 ThrIleArgLysAsnSerArgIleSerArgLeuGluAspTytrIleLysGluTytrCysAsn 102
Db 6786 AAAGCAGAGATGAGTAACGCTTTTCAAAATATTCAAGATTACATATTTGAATATTGTAAC 6845
QY 103 PheGlyCysAlaTytrPheValLeu----- 110
Db 6846 TTTGGTTGTCCTTACCTTTGTTATTAAAGACATTTGCTTAACAGAGGTTAAAGATTAAATC 6905
QY 111 GluAsnProAsnProArgAspIleLysPheAspGluArg 124
Db 6906 GAAGAAAGTTTCATCCGTCGTGATGATAGATGACGACAATAAA 6947
RESULT 13
ABQ69245/c
ID ABQ69245 standard; DNA; 3011208 BP.
XX
XX AC ABQ69245;
XX
XX 29-AUG-2002 (first entry)
XX
XX DE Listeria innocua DNA sequence #684.
XX
XX KW Antibacterial; Listeria; food contamination; mutational analysis;
XX infection; ds.
XX
XX OS Listeria innocua.
XX
XX PN WO200228891-A2.
XX
XX PD 11-APR-2002.
XX
XX PF 04-OCT-2001; 2001WO-FR03061.
XX
XX PR 04-OCT-2000; 2000FR-0012697.
XX
XX PA (INSP) INST PASTEUR.
XX (CNRS) CNRS CENT NAT RECH SCI.
XX
XX PI Kunst F, Glaser P;
XX WPI; 2002-332479/37.
XX
XX PT New genomic sequences from Listeria species, useful for detection,
XX treatment and prevention of infection, also related polypeptides,
XX antibodies and modulators -
XX
XX PS Clalim 5; SEQ ID 2058; 180pp; French.
XX
XX CC The present invention relates to nucleic acid sequences

CC containing them which are used to produce hosts cells which express the
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
 CC used to vaccinate subjects and to raise antibodies against the bacteria.
 CC The polypeptides may also be used to assay for other inhibitors of their
 CC activity and therefore identify compounds that may be used for the
 CC treatment of *S. epidermidis* infections, e.g. endocarditis. AAH53971 to
 CC AAH55090 represent specifically claimed *S. epidermidis* genomic DNA
 CC polynucleotide sequences from the present invention. AAH55091 to
 CC AAH55098 represent oligonucleotide sequences and primers which are used
 CC in the exemplification of the present invention.
 CC N.B. The present invention specifically claims all the polynucleotide
 CC sequences given in the sequence listing of the present specification,
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
 CC no sequences are present for SEQ ID NO:4455 to 4464.

XX SQ Sequence 3054 BP; 955 A; 532 C; 440 G; 1127 T; 0 other;

Alignment Scores:
 Pred. No.: 8.17e-19 Length: 3054
 Score: 255.50 Matches: 48
 Percent Similarity: 68.22% Conservative: 25
 Best Local Similarity: 44.86% Mismatches: 31
 Query Match: 25.32% Indels: 3
 DB: 22 Gaps: 2

US-09-868-352-23 (1-192) x AAH54843 (1-3054)

Qy 20 PheIleHisPheGluAsnIleValLysSerAspIleGluPheGlnLeuValIleAsn 39
 Db 2328 TTTTGTAAAGTGAAATATGATTAAAGTCGACCAACAATATTTTGAATGATAGAA 2269
 Qy 40 GluLysSerAlaPheAspValThrValPheGlyGlnArgPheSerGluIleLeuLys 59
 Db 2268 TATAGACAACTTTTGTATGACGAATATTTTCACCTAGGTATTCGGATATATAGACAAA 2209
 Qy 60 TyrAspPheIleValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLys 79
 Db 2208 TATGATTATGTCGTAGTACTATGTTACGATCAATTACGCTTAAAGGATTTTATATA 2149
 Qy 80 AsplaserThrIleArgLys--AsnSerArgIleSerArgLeuGluAspTyrIleLys 98
 Db 2148 GATAGTAAATAAAAGCGCAAAATAAGTAAACGATTTTCAAGTATACAAAGATATATACTA 2089
 Qy 99 GluTyrCysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspIle 118
 Db 2088 GAATATTGTAATTTGGTGTCTTATTTGTCTACGACGATTGTCACCAATGAATTT 2029
 Qy 119 -----LysPheAspAspGlu 123
 Db 2028 ATTTGAAGAAATAGATGATAA 2008

RESULT 11

ABN92594

ID ABN92594 standard; DNA; 399 BP.

XX AC ABN92594;

XX ABN92594;

XX 24-JUL-2002 (first entry)

XX Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:2057.

DE Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;

XX antibacterial; gene therapy; gene; ds.

XX Staphylococcus epidermidis.

OS US6380370-B1.

PN 30-APR-2002.

XX 13-AUG-1998; 98US-0134001.

XX

PR 14-AUG-1997; 97US-055779p.
 PR 08-NOV-1997; 97US-064964p.
 XX (GENO-) GENOME THERAPEUTICS CORP.
 PA Doucette-Stamm LA, Bush D;
 XX WPI; 2002-381255/41.
 PI P-PSDB; ABP40049.
 XX Novel isolated nucleic acid encoding a *Staphylococcus epidermidis*
 PT polypeptide, useful for diagnosing and treating bacterial infections -
 XX Disclosure; SEQ ID 2057; 267pp; English.
 PS ABN90538 to ABN93374 represent *Staphylococcus epidermidis* open reading
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
 CC given in ABP3124 to ABP37960. The *S. epidermidis* sequences have
 CC antibacterial activity and can be used in gene therapy. The sequences
 CC can also be used in the diagnosis and treatment of bacterial infections,
 CC particularly *S. epidermidis* infections. The sequences can be used to
 CC screen for compounds able to interfere with the *S. epidermidis* life
 CC cycle or inhibit *S. epidermidis* infection.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC USPTO web site.
 XX SQ Sequence 399 BP; 162 A; 41 C; 70 G; 126 T; 0 other;

Alignment Scores:
 Pred. No.: 7.24e-20 Length: 399
 Score: 254.50 Matches: 47
 Percent Similarity: 68.93% Conservative: 24
 Best Local Similarity: 45.63% Mismatches: 29
 Query Match: 25.22% Indels: 3
 DB: 24 Gaps: 2

US-09-868-352-23 (1-192) x ABN92594 (1-399)

Qy 24 GluAsnIleValLysSerAspIleGluPheGlnLeuValIleAsnGluLysSerAla 43
 Db 4 GAAATATGATGTTAAAGTCGACCAACAATATTTTGAATGATAGAAATGT 63
 Qy 44 PheAspValThrValPheGlyGlnArgPheSerGluIleLeuLysTyrAspPheIle 63
 Db 64 TTTGATCAGGAATATTTTCAGCTAGGTATTCGGATATATAGACAAATATGATTATGTC 123
 Qy 64 ValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAspAlaSerThr 83
 Db 124 GTAGGTGACTATGCTTACGATCAATTAACGCTTAAAGGATTTTATAAGATAGTAAATA 183
 Qy 84 IleArgLys--AsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyrCysAsn 102
 Db 184 AAGCGAGAAATAAGTAAACGATTTTCAAGTATATACAGATTATATAGTAAATTTGTAAT 243
 Qy 103 PheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspIle-----LysPhe 120
 Db 244 TTTGGTGTCTTATTTTGTAGTCAGACGATTGTCACCAATGAATTTTGTGAAGAAATA 303
 Qy 121 AspAspGlu 123
 Db 304 GATGATAAA 312

RESULT 12

AAV74525

ID AAV74525 standard; DNA; 7769 BP.

XX AC AAV74525;

XX 16-MAR-1999 (first entry)

XX Staphylococcus aureus contig SEQ ID #214.

XX

CC from the present invention can be used in diagnosis kits and assays, and
 CC pharmaceutical compositions and vaccines for *S. pneumoniae*.
 XX
 SQ Sequence 5558 BP; 1628 A; 1239 C; 1067 G; 1624 T; 0 other;

Alignment Scores:
 Pred. No.: 3.36e-39 Length: 5558
 Score: 438.00 Matches: 88
 Percent Similarity: 68.89% Conservativity: 36
 Best Local Similarity: 48.89% Mismatches: 50
 Query Match: 43.41% Indels: 6
 DB: 19 Gaps: 4

US-09-868-352-23 (1-192) x AAV52236 (1-5558)

Qy 1 MetArgLysGluValThrProGluMetLeuAsnTyrAsnLysTyrProGlyProGlnPhe 20
 Db 4885 ATGCGAAAGAAATGACCTGAATTACAACTATAACAAGTTTCCTGGTCCGGAGTTC 4826
 Qy 21 IleHisPheGluAsnIleValLysSerAspIleGluPheGlnLeuValIleAsnGlu 40
 Db 4825 CATTTACACGGGGACAGCTCGAACGGAAGGATAGCTTTTCTTGGTTGAAATATC 4766
 Qy 41 LysSerAlaPheAspValThrValPheGlyGlnArgPheSerGluIleLeuLysTyr 60
 Db 4765 AAGGATGCTTTGATGTGACGACTTTTAATCAGCGCTTTTTCAGAAATTAACCAAGTTT 4706
 Qy 61 AspPheIleValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAsp 80
 Db 4705 GATTATATCGTGGGGAGCTGGAGCAACCAACAGCTTCGCCTACGAGGTTTTTACAAGGAT 4646
 Qy 81 AlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyr 100
 Db 4645 GACCGAACGAAAGAAACTTGAATAAATCAGTCGTTTACAGAGACTACCTTTTACAGTAT 4586
 Qy 101 CysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspIleLysPhe 120
 Db 4585 TGTAGTTATGTTGTCCTATTTGCTAGAAAATGAAGCCCTAAGCGAGCATCATTT 4526
 Qy 121 AspAspGluArgProHisLysArgLysSerArgSerLysSerGlnSerLysSer 140
 Db 4525 GAC-----AAGAAATCGTGAACGGAAGAAACACCTCTAGAAAAGGAAAGAAA 4472
 Qy 141 GlnThrArgAsnArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysLysArg 160
 Db 4471 CCGACTCAAACTAAACGA---AAGTCGAATCGAGATAGAAA---AATAGACGTCGTCAG 4418
 Qy 161 LysAspThrLysArgArgGlnGluArgHisIleLysGlu-----GlnAspLysGlu 178
 Db 4417 AAGACGACAAAGGACAAACGTCATTTGCTCATTCGTCAGAAAGTCAGTAGAGAAATAGGAG 4358
 RESULT 8
 ID ABA90521 standard; DNA; 2365589 BP.
 XX
 AC ABA90521;
 XX
 DT 16-MAY-2002 (first entry)
 XX
 DE Genomic sequence of *Lactococcus lactis* IL1403.
 XX
 KW Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese; ds.
 XX
 OS *Lactococcus lactis* IL1403.
 XX
 PN FR2807446-A1.
 XX
 XX 12-OCT-2001.
 XX
 PD 11-APR-2000; 2000FR-0004630.
 XX
 PF 11-APR-2000; 2000FR-0004630.
 XX
 PR 11-APR-2000; 2000FR-0004630.
 XX

PA (INRG) INRA INST NAT RECH AGRONOMIQUE.
 XX
 PI Bolotine A, Sorokine A, Renault P, Ehrlich SD;
 XX
 DR WPI; 2002-043418/06.
 XX
 PT New nucleotide sequence useful in the identification of *Lactococcus*
 XX
 PS Claim 1; SEQ ID 1; 2504pp; French.
 XX
 CC The present invention is related to a *Lactococcus lactis* nucleotide
 CC sequence (ABA90521) and related proteins (ABB53300-ABB55621). The
 CC nucleic acid sequence is useful in the detection and/or amplification of
 CC nucleic acid sequence, particularly to identify *Lactococcus lactis* or
 CC related species. The proteins of the invention are useful for the
 CC biosynthesis or biodegradation of a composition of interest. The
 CC invention helps research in lactic bacteria, particularly useful in the
 CC production of yogurt and cheese.
 CC Note: The sequence data for this patent is based on equivalent patent
 CC WO200177334 (published 18-OCT-2001) which is available in electronic
 CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 2365589 BP; 765914 A; 415261 C; 420487 G; 763927 T; 0 other;

Alignment Scores:
 Pred. No.: 1.24e-23 Length: 2365589
 Score: 331.50 Matches: 75
 Percent Similarity: 51.83% Conservativity: 38
 Best Local Similarity: 34.40% Mismatches: 64
 Query Match: 32.85% Indels: 41
 DB: 24 Gaps: 5

US-09-868-352-23 (1-192) x ABA90521 (1-2365589)

Qy 1 MetArgLysGluValThrProGluMetLeuAsnTyrAsnLysTyrProGlyProGlnPhe 20
 Db 183231 ATGCTAAAGTAATAGACGAAAGTAAGTTAATAATAATAATCTCGGCAACATGTT 183290
 Qy 21 IleHisPheGluAsnIleValLysSerAspIleGluPheGlnLeuValIleAsnGlu 40
 Db 183291 ATGCGAGTGGGGAGTGTTCACAGTTGGACACGAACTTTTCATATTTGTCATAAATAT 183350
 Qy 41 LysSerAlaPheAspValThrValPheGlyGlnArgPheSerGluIleLeuLysTyr 60
 Db 183351 CGTGAGGCTTTTGTATGCGGAAAAATTAGAACACAGCTTTTCTGATGTTTAGATAAGTAT 183410
 Qy 61 AspPheIleValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAsp 80
 Db 183411 GATTATATTGCGGAGATGGGCTTTCAGCAATTCGCTTGAAGGGTTTTTTC----- 183464
 Qy 81 AlaSerThrIleArgLys-----AsnSerArgIleSerArgLeuGluAspTyr 96
 Db 183465 ---TCAACCCAGCTCGTAGAAAAATGTTGGCAGACATAAAATTCATCATCTTGAAGATTAT 183521
 Qy 97 IleLysGluTyrCysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArg 116
 Db 183522 GTCAATGAATATTGTAATTACGGTTTGCTTATTGTTTACCGCGGATTCGGACAAAA 183581
 Qy 117 AspIleLysPheAspAspGluArg-----
 Db 183582 GATCAAGCTTTTGTCTTGAGAAACTTTTACCGAAAAAGAGTTCAACCAAGGATTTGAT 183641
 Qy 125 -----ProHisLysArgArgLysSerSerLysSerGlnSerSerLysSerGlnThr 142
 Db 183642 AAGCCACGTCGTAAAGAGAAAATCGTAATCGAAATTTGGCGCTGATGAACACAAAATTTACT 183701
 Qy 143 ArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysLysArgLysAsp 162
 Db 183702 AAGAAGACAAACCGTTTCGGAAT-----TCATCAGAAAGCTCGAANAGAT 183746
 Qy 163 ThrLysArgArgGlnGluArgHis-----
 ||| |||:||||:

PR 27-NOV-2000; 2000US-253625P.
 PR 22-DEC-2000; 2000US-257931P.
 PR 16-FEB-2001; 2001US-269308P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 XX
 XX WPI; 2001-611495/70.
 DR P-PSDB; AAU37973.
 XX
 PT New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 XX
 XX Claim 27; Seq ID No 9469; 511pp; English.
 XX
 CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence encodes an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 531 BP; 189 A; 93 C; 118 G; 131 T; 0 other:

Alignment Scores:
 Pred. No.: 3,42e-42 Length: 531
 Score: 452.50 Matches: 94
 Percent Similarity: 56.49% Conservative: 33
 Best Local Similarity: 49.21% Mismatches: 48
 Query Match: 44.85% Indels: 16
 DB: 23 Gaps: 4

US-09-868-352-23 (1-192) x AAS55832 (1-531)

QY 1 MetArgLysGluValThrProGluMetLeuAsnTyrAsnLysTyrProGlyProGlnPhe 20
 DB 1 ATCGAAAGAAATGTCACCTGAATTATACAACTATATACAAAGTTTCTGCTCGGAGTTC 60
 QY 21 IleHisPheGluAsnIleValLysSerAspAspIleGluPheGlnLeuValIleAsnGlu 40
 DB 61 CATTTACCGGGGACAAAGGTCGAAACAGGAGGATAGCTTTTCCCTGGTGTGAAATATC 120
 QY 41 LysSerAlaPheAspValThrValPheGlyGlnArgPheSerGluIleLeuLysTyr 60
 DB 121 AAGGATCGCTTGTATGTGACGACTTTTAAATCAGGTTTTTTCAGAACTTAACCAAGTTT 180
 QY 61 AspPheIleValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAsp 80
 DB 181 GATTATATCGTGGGGGACTGGACGACGACAGCTTCGCTCAGGAGGATTTTACAGGAT 240
 QY 81 AlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyr 100
 DB 241 GACCGAACAGAGAAACCTTGAAATAATCAGTCGTTTACAGAACTACCTTTTAGAGTAT 300
 QY 101 CysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspIleLysPhe 120
 DB 301 TGTAGTTATGTTGTGCTATTGTTGCTTAGAAATGAAGCCCTTAAGCGAGCATCATTT 360

QY 121 AspAspGluArgProHisLysArgArgLysSerArgSerLysSerGlnSerSerLysSer 140
 DB 361 GAC-----AAGAAATATGCGTAAGACGGAAGAAACACCTTCTAGAAAAGGAAGAAA 414
 QY 141 GlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysLysArg 160
 DB 415 CCACACTCAAACTAAACGA---AAGTCGAATGCAGAT----- 447
 QY 161 LysAspThrLysArgArgGluGluArgHisIleLysGluGlnAspLysGluMetThr 180
 DB 448 AAGAAAAATAGACGTCGTCAGAAAGACCACCATCTCAGAAAGAGGACAGGAA----- 501
 QY 181 SerAlaLysGlnHis-IeuLeuPheValArg 190
 DB 502 -----CAACGTCATTTTGTGTCATTCGTCAGA 526
 RESULT 7
 AAV52236/c
 ID AAV52236 standard; DNA; 5558 BP.
 XX
 AC AAV52236;
 DT 23-OCT-1998 (first entry)
 XX
 DE Streptococcus pneumoniae genome fragment SEQ ID NO:103.
 XX
 KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
 KW computer readable medium; vaccine; pharmaceutical composition; ds.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN W09818931-A2.
 XX
 PD 07-MAY-1998.
 XX
 PF 30-OCT-1997; 97WO-US19588.
 XX
 PR 31-OCT-1996; 96US-0029960.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;
 PI Kunsch CA, Rosen CA;
 XX
 DR WPI; 1998-272225/24.
 XX
 PT Computer-readable medium with recorded Streptococcus pneumoniae
 PT polynucleotide sequences - useful in diagnostic kits and assays, and
 PT pharmaceutical compositions and vaccines for Streptococcus
 PT pneumoniae
 XX
 PS Claim 1; Page 770-773; 1409pp; English.
 XX
 CC The present invention describes a computer readable medium which has
 CC the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)
 CC recorded on it, or a representative fragment or a sequence at least 95%
 CC identical to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in
 CC SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from
 CC Streptococcus pneumoniae. The present invention also describes an
 CC isolated nucleic acid molecule encoding a homologue of any of the
 CC fragments of the S.pneumoniae genome (SEQ ID NO:1 to 391) where the
 CC nucleic acid molecule is produced by a process comprising: (a) screening
 CC a genomic DNA library using as a probe a target sequence defined by any
 CC of the sequences in SEQ ID NO:1 to 391, identifying members of the
 CC library which contain sequences that hybridise to the target sequence and
 CC isolating the nucleic acid molecules from the members; or (b) isolating
 CC mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid
 CC molecules whose nucleotide sequence is homologous to amplification
 CC primers derived from the fragment of the S. pneumoniae genome to prime
 CC the amplification and isolating the amplified sequences. The computer
 CC readable medium can be used in a computer-based system for identifying
 CC fragments of the S. pneumoniae genome of commercial importance, or
 CC expression modulating fragments of the S. pneumoniae genome. Products

```

Db 562 CATAAAGAAAGTCA---AAACGTAATCAGACTAGTCAACTTAATACCAAAATTTAGTCAT 618
QY 186 LeuLeuPheValArgLysAsn 192
Db 619 TTTATCATCAGAAAGAAAGAT 639

RESULT 5
AAS55679
ID AAS55679 standard; DNA; 531 BP.
AC AAS55679;
XX
XX
XX 13-FEB-2002 (first entry)
XX
DE Streptococcus pneumoniae DNA for cellular proliferation protein #250.
XX
XX Antisense; ds; prokaryotic cellular proliferation gene;
KW antibiotic; antibacterial; drug design.
XX
OS Streptococcus pneumoniae.
XX
PN WO200170955-A2.
XX
XX 27-SEP-2001.
XX
XX 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
XX WPI; 2001-611495/70.
DR P-PSDB; AAU37820.
XX
XX New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids -
PS Claim 27; Seq ID No 9316; 511pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
XX prokaryotic cellular proliferation, their use in identifying the
XX genes, their use in the discovery of novel antibiotics, the essential
XX genes themselves and the encoded proteins. The prokaryotes used are
XX Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
XX invention is also useful for the identification of potential new targets
XX for antibiotic development. The antisense nucleic acids can also be used
XX to identify proteins used in proliferation, to express these proteins,
XX and to obtain antibodies capable of binding to the expressed proteins.
XX The proteins can be used to screen compounds in rational drug discovery
XX programmes. The antisense nucleic acid sequence is also useful to screen
XX for homologous nucleic acids which are required for cell proliferation in
XX a wide variety of organisms. The present sequence encodes an
XX essential prokaryotic cellular proliferation protein.
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 531 BP; 190 A; 92 C; 118 G; 131 T; 0 other;
Alignment Scores:
Pred. No.: 3.42e-42 Length: 531

```

```

Score: 452.50 Matches: 94
Percent Similarity: 66.49% Conservative: 33
Best Local Similarity: 49.21% Mismatches: 48
Query Match: 44.85% Indels: 16
DB: 23 Gaps: 4

US-09-868-352-23 (1-192) x AAS55679 (1-531)
QY 1 MetArgLysGluValThrProGluMetLeuAsnTyrAsnLysTyrProGlyProGlnPhe 20
Db 1 ATGCGAAAAGAAATTCACCTGAATTATACAACTATAACAGTTTCCTGGTCCGAGATC 60
QY 21 IleHisPheGluAsnIleValLysSerAspIleGluPheGlnLeuValIleAsnGlu 40
Db 61 CATTTACACGGGACAGGTGGAACGGAAGGATAGCTTTTTCCTTGGTTGAAATATC 120
QY 41 LysSerAlaPheAspValThrValPheGlyGlnArgPheSerGluIleLeuLysTyr 60
Db 121 AAGGATGCTTTGATGTGACGACTTTTAAATCAGCGCTTTTTCAGAAAGTATTAACCAAGTTT 180
QY 61 AspPheIleValGlyAspTrpClyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAsp 80
Db 181 GATTATATCGTGGGGACTGGAGCAACGACAGCTTCGCCCTACGAGGTTTTTACAGGAT 240
QY 81 AlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyr 100
Db 241 GACCGAACAGAGAAACACTTGAAAAAATCAGTCGCTTTTACAGACTACTCTTTAGAGTAT 300
QY 101 CysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspIleLysPhe 120
Db 301 TGAGTTATGGTTGTGCTATTTTCTAGAAAAATGAAGCCCTTAAGCGAGCATCATTT 360
QY 121 AspAspGluArgProHisLysArgArgLysSerArgSerLysSerGlnSerSerLysSer 140
Db 361 GAC-----AAGAAATCGGTAAGACGGAAGAACACCTTCTAGAAAAAGGAAGAAA 414
QY 141 GlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysLysArg 160
Db 415 CCGACTCAAACTAAACGA---AAGTCGATGCAGAT----- 447
QY 161 LysAspThrLysArgArgGlnGlnArgHisIleLysGluGlnAspLysGluMetThr 180
Db 448 AAGAAAAATAGACGTCGTCAGAAAGACCGACGATTTCTCAGAAAGAGACAAAGGAA----- 501
QY 181 SerAlaLysGlnHis-LeuLeuPheValArg 190
Db 502 -----CAACGTCATTTTGTTCATTGTCGACAG 526

RESULT 6
AAS55832
ID AAS55832 standard; DNA; 531 BP.
XX
XX AC AAS55832;
XX
XX 13-FEB-2002 (first entry)
XX
DE Streptococcus pneumoniae DNA for cellular proliferation protein #403.
XX
XX Antisense; ds; prokaryotic cellular proliferation gene;
KW antibiotic; antibacterial; drug design.
XX
OS Streptococcus pneumoniae.
XX
PN WO200170955-A2.
XX
XX 27-SEP-2001.
XX
XX 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.

```

Db 1 ATCGGAAAGAAAGTGACACAGAGATGCTTAACATATAAAGTATCTCTGCCCCACAGTTC 60
 QY 21 IleHisPheGluAsnIleValIysSerAspIleGluPheGlnLeuValIleAsnGlu 40
 Db 61 ATTCACATTTCAAAATATCGTTAAAGTGATGATATTGAATTTCAACTTGTATTAAATGAA 120
 QY 41 LysSerAlaPheAspValThrValPheGlyClnArgPheSerGluIleLeuLeuLysTyr 60
 Db 121 AAATACAGCTTTTGATGATGCTGCTTTGGACAACGTTTCTCAGATTTTATTAAATAT 180
 QY 61 AspPheIleValGlyAspTTPGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAsp 80
 Db 181 GATTTATCGTTGCGGATTCGGGTACGACAGCTTGAGGCTAGAGCGCTTTTCAAGAT 240
 QY 81 AlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyr 100
 Db 241 GCTAGTACAATTAGAAAAATAGCCGATTTACGTTTACAAGATTATATAAAGAGTAT 300
 QY 101 CysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspIleLysPhe 120
 Db 301 TGTAACTTTTGGTGTGCTTATTGTTGTTGGAGAATCCAAATCCTAGAGATATTAAATTT 360
 QY 121 AspAspGluArgProHisLysArgLysSerArgLysSerLysSerLysSerLysSer 140
 Db 361 GATGATCAAGACCTCATAGCGCTGTAAGTCAAGATCCAAATCAACAATCATCAAGTCA 420
 QY 141 GlnThrArgAsnAsnArgSerGlnSerAsnAlaHisPheThrSerLysLysArg 160
 Db 421 CAAACTAGAAATAATCGTCCAGTCAAAATGCCAATGCTCATTTTACAAGTAAAGCGT 480
 QY 161 LysAspThrLysArgArgGlnGluArgHisIleLysGluGlnAspLysGluMetThr 180
 Db 481 AAAGACACAAAGCCGCTCAAGAACGTCATATTAAAGAGCAAGCAAGATAAGGAATGACC 540
 QY 181 SerAlaLysGlnHis-LeuLeuPheValArgLys 191
 Db 541 TCTGCAAGCAGCATTTGTTATTCGTAGAAAA 574
 RESULT 4
 ID ABN66118 standard; DNA: 642 BP.
 AC ABN66118;
 XX
 XX 01-JUL-2002 (first entry)
 DE Streptococcus polynucleotide SEQ ID NO 149.
 KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
 KW anti-inflammatory; infection; vaccine; meningitis; gene therapy; ds.
 XX
 OS Streptococcus pyogenes.
 XX
 XX WO200234771-A2.
 XX
 XX 02-MAY-2002.
 XX
 XX 29-OCT-2001; 2001WO-GB04789.
 XX
 XX 27-OCT-2000; 2000GB-0026333.
 XX
 XX 24-NOV-2000; 2000GB-0028727.
 XX
 XX 07-MAR-2001; 2001GB-0005640.
 XX
 XX (CHIR-) CHIRON SPA.
 XX (GENO-) INST GENOMIC RES.
 XX
 XX Telford J, Massignani V, Margarit Ros YI, Grandi G, Fraser C;
 XX Tettelin H;
 XX
 XX WPI; 2002-352536/38.
 XX
 XX P-PSDB; ABP25487.

PT New Streptococcus protein for the treatment or prevention of infection
 PT or disease caused by Streptococcus bacteria, such as meningitis, and
 PT for detecting a compound that binds to the protein -
 XX
 PS Claim 7; Page 3168; 4525pp; English.
 XX
 CC The invention relates to a protein (ABP25413-ABP30895) from group B
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
 CC the specification. The proteins have antibacterial and anti-inflammatory
 CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
 CC antibodies that bind (I) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a
 CC biological sample. (I) is used to determine whether a compound binds to
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (I) may be used to recombinantly produce (I) and may be
 CC used in gene therapy. Antibodies to (I) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins.
 XX
 XX Sequence 642 BP; 251 A; 106 C; 114 G; 171 T; 0 other;
 SQ
 Alignment Scores:
 Pred. No.: 4, 01e-54 Length: 642
 Score: 558.50 Matches: 107
 Percent Similarity: 71.98% Conservative: 42
 Best Local Similarity: 51.69% Mismatches: 41
 Query Match: 55.35% Indels: 17
 DB: 24 Gaps: 4
 US-09-868-352-23 (1-192) x ABN66118 (1-642)
 QY 1 MetArgLysGluValThrProGluMetLeuAsnTyrAsnLysTyrProGlyProGlnPhe 20
 Db 25 ATGAAAAAGAAAAATTTGCCGAGAAATGTACACTATAATAATTTCCAGGTCCTCAAAATTT 84
 QY 21 IleHisPheGluAsnIleValLysSerAspIleGluPheGlnLeuValIleAsnGlu 40
 Db 85 ATTCAATTTGAAGAGCAAGTTAAGGCTGAAGCATTTGTTACTCTTAGAAGATGTT 144
 QY 41 LysSerAlaPheAspValThrValPheGlyClnArgPheSerGluIleLeuLysTyr 60
 Db 145 AAGAAGCGCTTTTGATACGACTAGCTTTGGTCAAGCTTATACAGAAGTCTCTCTAAAGTAT 204
 QY 61 AspPheIleValGlyAspTTPGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAsp 80
 Db 205 GACTATATTGTCGGATTCGGGAAATGAACAGCTTCGTCTTAAAGGGTTTATAAGGAT 264
 QY 81 AlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyr 100
 Db 265 AGTCACCATATCAAGAAAAACGAATCGCATCTCAGTTTAGAAGATTATATAAAGAATTT 324
 QY 101 CysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspIleLysPhe 120
 Db 325 TGCAATTTTGGTGTGCTTATTGTTTGAAGAAATCTTCATCCACAAGATATTAAATTT 384
 QY 121 AspAspGluArgProHisLysArgLysSer---ArgSerLysSerGln----- 136
 Db 385 GAAGAGGAGCGCCCAACCAAGACGAAAGAAATCACCTAAATCAAAATCAATTCGTCGAAG 444
 QY 137 -----SerSerLysSerGlnThrArgAsnAsn 145
 Db 445 CCAAACTATTTCAATCAGACGCCGCCAACACCTAAGAGCAAAATCGAAG---CGGGCATCA 501
 QY 146 ArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysLysArgLysAspThrLysArg 165
 Db 502 AAAGAAAAGCAACCTGAACCAACCAAGCGCTTACCAGTCAAAACGTCGAAGTAATACTAAG 561
 QY 166 ArgGlnGluArgHisIleLysGluGlnAspLysGluMetThrSerAlaLysGlnHis 185

```
XX
PS Claim 8; Page 4196-4488; 4525pp; English.
CC The invention relates to a protein (ABP25413-ABP30895) from group B
CC Streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins.
SQ Sequence 2155561 BP; 695741 A; 385320 C; 382958 G; 691542 T; 0 other;

Alignment Scores:
Pred. No.: 3.65e-99 Length: 2155561
Score: 996.00 Matches: 192
Percent Similarity: 99.48% Conservative: 0
Best Local Similarity: 99.48% Mismatches: 0
Query Match: 98.71% Indels: 1
DB: 24 Gaps: 0

US-09-868-352-23 (1-192) x ABN71527 (1-2155561)
QY 1 MetArgLysGluValThrProGluMetLeuAsnTyrAsnLysTyrProGlyProGlnPhe 20
Db 442722 ATGCCAAAAGAGTACACACAGAGATGCTTAACATAATAAGTATCCCGCCACAGTTC 442781
QY 21 IleHisPheGluAsnIleValLysSerAspIleGluPheGlnLeuValIleAsnGlu 40
Db 442782 ATTCACTTTGAAAATATCGTTAAAGATGATGATATGAATTCACATGTTATTAAAGAA 442841
QY 41 LysSerAlaPheAspValThrValPheGlyGlnArgPheSerGluIleLeuLysTyr 60
Db 442842 AAATCAGCTTTTGATGTGACTGTCTTTGGACAAACGTTTCTGAGATTTTATAAAATAT 442901
QY 61 AspPheIleValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAsp 80
Db 442902 GATTTTATCGTTGGCGATTGGGGTAACGAGCAGTTGAGCGCTTAAGAGGCTTTTACAAAGAT 442961
QY 81 AlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyr 100
Db 442962 GCTAGTACAATTAGAAAAAATAGCGGATTTTCACGTTTAGAAGATTTATTTAAAGAGTAT 443021
QY 101 CysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspIleLysPhe 120
Db 443022 TGTAACTTTGGTGTGCTTATTTGTGTGGAGAATCCAAATCCTAGAGATATTAAATTT 443081
QY 121 AspAspGluArgProHisLysArgArgLysSerArgSerLysSerGlnSerSerLysSer 140
Db 443082 GATGATGAAGAACCTTCATAGCGTGTAGTCAAGATCCAATCACAATCATCAAGTCA 443141
QY 141 GlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysLysArg 160
Db 443142 CAAACTAGAAATAATCGTTCACAGTCAAAATGCCAATGCTCATTTTACAAGTAAAGCGT 443201
QY 161 LysAspThrLysArgArgGlnGluArgHisLysLysGluGlnAspLysGluMetThr 180
Db 443202 AAAACACAAAACCGCGTCAAGACGTCATATTAAAGAAGAGCAAGATTAAGAAATGACC 443261
QY 181 SerAlaLysGlnHis-LeuLeuPheValArgLysAsn 192
Db 443262 TCTGCAAGACGACATTTTGTATTATTCGTGAAGAAAAAAT 443298
```

RESULT 3

```
ABN66117
ID ABN66117 standard; DNA; 576 BP.
XX
XX ABN66117;
XX AC
XX DT 01-JUL-2002 (first entry)
XX DE Streptococcus polynucleotide SEQ ID NO 147.
XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
XX group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
XX antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX OS Streptococcus agalactiae.
XX PN WO200234771-A2.
XX PD 02-MAY-2002.
XX PF 29-OCT-2001; 2001WO-GB04789.
XX PR 27-OCT-2000; 2000GB-0026333.
XX PR 24-NOV-2000; 2000GB-0028727.
XX PR 07-MAR-2001; 2001GB-0005640.
XX PA (CHIR-) CHIRON SPA.
XX PA (GENO-) INST GENOMIC RES.
XX PI Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;
XX Tettelin H;
XX WI: 2002-352536/38.
XX P-PSDB: ABP25486.
XX PT New Streptococcus protein for the treatment or prevention of infection
XX or disease caused by Streptococcus bacteria, such as meningitis, and
XX for detecting a compound that binds to the protein -
XX PS Claim 7; Page 3168; 4525pp; English.
XX CC The invention relates to a protein (ABP25413-ABP30895) from group B
XX streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
XX (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
XX the specification. The proteins have antibacterial and antiinflammatory
XX activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
XX antibodies that bind (I) are used in the manufacture of medicaments for
XX the treatment or prevention of infection or disease caused by
XX Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
XX Nucleic acids encoding (I) are used to detect Streptococcus in a
XX biological sample. (I) is used to determine whether a compound binds to
XX (I). A composition comprising (I) or a nucleic acid encoding (I), may be
XX used as a vaccine or diagnostic composition. The disease caused by
XX Streptococcus that is prevented or treated may be meningitis. Nucleic
XX acid encoding (I) may be used to recombinantly produce (I) and may be
XX used in gene therapy. Antibodies to (I) are used for affinity
XX chromatography, immunoassays, and distinguishing/identifying
XX Streptococcus proteins.
SQ Sequence 576 BP; 212 A; 86 C; 108 G; 170 T; 0 other;

Alignment Scores:
Pred. No.: 3.38e-103 Length: 576
Score: 990.00 Matches: 191
Percent Similarity: 99.48% Conservative: 0
Best Local Similarity: 99.48% Mismatches: 0
Query Match: 98.12% Indels: 1
DB: 24 Gaps: 0

US-09-868-352-23 (1-192) x ABN66117 (1-576)
QY 1 MetArgLysGluValThrProGluMetLeuAsnTyrAsnLysTyrProGlyProGlnPhe 20
|||||
```

/product= N-acetyl_muramidase_homologue

FT XX PN WO200037646-A2.
 XX PD 29-JUN-2000.
 XX PF 22-DEC-1999; 99WO-GB04377.
 XX PR 22-DEC-1998; 98GB-0028345.
 XX PR 22-DEC-1998; 98GB-0028349.
 XX PR 22-DEC-1998; 98GB-0028350.
 XX PR 22-DEC-1998; 98GB-0028352.
 XX PR 22-DEC-1998; 98GB-0028353.
 XX PR 22-DEC-1998; 98GB-0028354.
 XX PR 22-DEC-1998; 98GB-0028355.
 XX PR 22-DEC-1998; 98GB-0028356.
 XX PR 22-DEC-1998; 98GB-0028357.
 XX PR 22-DEC-1998; 98GB-0028359.
 XX PR 04-JAN-1999; 99GB-0000082.
 XX PR 04-JAN-1999; 99GB-0000083.
 XX PR 04-JAN-1999; 99GB-0000084.
 XX PR 04-JAN-1999; 99GB-0000085.
 XX PR 28-JAN-1999; 99GB-0000086.
 XX PR 28-JAN-1999; 99GB-0001916.
 XX PR 28-JAN-1999; 99GB-0001922.
 XX PA (MICR-) MICROSCIENCE LTD.
 XX PI Hughes MJG, Santangelo JD, Lane JD, Everest P, Feldman R;
 XX PI Moore JC, Wilson RK, Dobson RJ, Dougan G;
 XX DR WPI; 2000-442674/38.
 XX P-PSDB; AAY96805.
 XX PT New peptides useful for treatment and prevention of conditions
 XX PT associated with Streptococcal infection are obtained from group B
 XX PT Streptococcus
 XX PS Example 11; Page 46-47; 63pp; English.
 XX CC The S. agalactiae pho3-1 gene encodes a protein which has homology with
 XX CC proteins of unknown function in S. pyogenes, S. pneumoniae,
 XX CC Bacillus subtilis (yutD) and Enterococcs faecillis. The B. subtilis yutD
 XX CC gene is located in a chromosomal region containing genes involved in
 XX CC cell wall synthesis.
 XX CC Peptides derived from group B Streptococcus (Streptococcus agalactiae)
 XX CC proteins encoded by genes pho1-13, pho3-21, pho2-15, pho3-18, pho3-22,
 XX CC pho3-3, pho3-17, pho2-2, pho1-5, pho3-1, pho3-23, pho3-50, pho1-14,
 XX CC pho2-10, pho3-14, pho3-24 and pho3-29. The peptides are useful for
 XX CC screening potential drugs, or for the detection of virulence, and for the
 XX CC manufacture of a medicament for use in the treatment or prevention of
 XX CC infections such as focal infection (including osteomyelitis, septic
 XX CC arthritis, abscesses and endophthalmitis) and urinary tract infections
 XX CC caused by group B Streptococcus. The peptides and vaccines comprising the
 XX CC peptides are useful in treatment of chronic mastitis, especially in cows
 XX CC (i.e. for veterinary purposes).
 XX SQ Sequence 579 BP; 213 A; 85 C; 109 G; 172 T; 0 other;

Alignment Scores:
 Pred. No.: 2,37e-105 Length: 579
 Score: 1009.00 Matches: 192
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 21 Gaps: 0

US-09-868-352-23 (1-192) x AAA51368 (1-579)

QY 1 MetArgLysGluValThrProGluMetLeuAsnTyrAsnLysTyrProGlyProGlnPhe 20
 Db 1 ATGCGAAAGAGTGACACCAGAGATGCTTAAGTATTAAGTATCTCTGCCCCACAGTTT 60

QY 21 IleHisPheGluAsnIleValLysSerAspIleGluPheGlnLeuValIleAsnGlu 40
 Db 61 ATTCACTTTCAAAATATCGTTAAAGCTGATGATATTGAATTTCAACTTGTATTATTAAGAA 120
 QY 41 LysSerAlaPheAspValThrValPheGlyGlnArgPheSerGluIleLeuLysTyr 60
 Db 121 AAATCAGCTTTTGATGTGACTGCTTTTGACACACGTTTTTCTGAGATTTTATTAAAAATAT 180
 QY 61 AspPheIleValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAsp 80
 Db 181 GATTTTATCGTTGGCGATTGGGGTAACGAGCAGTTCAGGCTAAGAGGCTTTTACAAAGAT 240
 QY 81 AlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyr 100
 Db 241 GCTAGTACGATTAGAAAAAATAGCCGGATTTCACGTTTAGAAGATTATATAAAGAGATAT 300
 QY 101 CysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspIleLysPhe 120
 Db 301 TGTAACTTTGGTTGCTTATTTTTGTGGAGAATCCAAATCCTAGAGATATTAAATTT 360
 QY 121 AspAspGluArgProHisLysArgArgLysSerArgSerLysSerGlnSerSerLysSer 140
 Db 361 GATGATGAAGACCTCATAAGCGTCGTAAAGTCAAGATCCAAATCACAATCATCAAGTCA 420
 QY 141 GlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysArg 160
 Db 421 CAAACTAGAAATAATCGTTCCCAATGCAATGCTCATTTTACAAGTAAAAAGCGT 480
 QY 161 LysAspThrLysArgArgGlnGluArgHisIleLysGluGlnAspLysGluMetThr 180
 Db 481 AAAGACACAAAACCCGCTCAAGAACGTCATATTAAAGAGAGCAGCAAGATAGGAAATGACC 540
 QY 181 SerAlaLysGlnHisLeuLeuPheValArgLysAsn 192
 Db 541 TCTGCAAGCAGCATTTGTTATTTCGTAAGAAAAAAT 576
 RESULT 2
 ABN71527
 ID ABN71527 standard; DNA; 2155561 BP.
 XX AC ABN71527;
 XX DT 02-JUL-2002 (first entry)
 XX DE Streptococcus polynucleotide SEQ ID NO 10967.
 XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 XX KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
 XX KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
 XX OS Streptococcus sp.
 XX PN WO200234771-A2.
 XX PD 02-MAY-2002.
 XX PF 29-OCT-2001; 2001WO-GB04789.
 XX PR 27-OCT-2000; 2000GB-0026333.
 XX PR 24-NOV-2000; 2000GB-0028727.
 XX PR 07-MAR-2001; 2001GB-0005640.
 XX PA (CHIR-) CHIRON SPA.
 XX PA (GENO-) INST GENOMIC RES.
 XX PI Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;
 XX PI Tettein H;
 XX DR WPI; 2002-352536/38.
 XX PT New Streptococcus protein for the treatment or prevention of infection
 XX PT or disease caused by Streptococcus bacteria, such as meningitis, and
 XX PT for detecting a compound that binds to the protein -

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 25, 2003, 09:34:32 ; Search time 227 Seconds
(without alignments)
1904.774 Million cell updates/sec

Title: US-09-868-352-23
Perfect score: 1009
Sequence: 1 MRKEVTPPEMLNKNYKPGPQF.....EEQDKEMTSKQHLFLVRKN 192

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+p2n.model -DEV=xlh
-Q/cgn2_1/USPTO.spool/US09868352/runat_14022003_153515_4212/app_query.fasta_1.391
-DB=N_Geneseq_101002 -QEXT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=40
-MODE=LOCAL -OUTFWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09868352 -CGN_1_187/runat_14022003_153515_4212 -NCPUS=6 -ICPU=3
-NO_XLPY -NO_MMAP -LARGESUB -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_101002.*
1: /SID52/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID52/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	
1	1009	100.0	579	21	AAA51368	Streptococcus agal	
2	996	98.7	2155561	24	ABN71527	Streptococcus poly	
3	990	98.1	576	24	ABN66117	Streptococcus poly	
4	558.5	55.4	642	24	ABN66118	Streptococcus poly	
5	452.5	44.8	531	23	AA555679	Streptococcus pneu	
6	452.5	44.8	531	23	AA555832	Streptococcus pneu	
c	7	438	43.4	5558	19	AAV52236	Streptococcus pneu
8	331.5	32.9	2365589	24	ABA90521	Genomic sequence o	
9	302.5	30.0	1495	20	AAH13439	Enterococcus faeca	
c	10	255.5	25.3	3054	22	AAH54843	S. epidermidis gen
11	254.5	25.2	399	24	ABN92594	Staphylococcus epi	
12	249.5	24.7	7769	18	AAV74525	Staphylococcus aur	
c	13	248	24.6	3011208	24	ABQ69245	Listeria innocua D
14	243.5	24.1	390	22	AAH52964	S. epidermidis ope	
c	15	240.5	23.8	2944528	24	ABA03041	Listeria monocytog
16	239.5	23.7	1234	24	ABQ68718	Listeria monocytog	
17	239.5	23.7	2690	24	ABQ70820	Listeria monocytog	
18	232.5	23.0	285	24	ABK80645	Bacillus clausii g	
19	222.5	22.1	273	24	ABK75854	Bacillus lichenifo	
20	113.5	11.2	6145	20	AAH84569	Biorhythm marker g	
21	108.5	10.8	1826	22	AAH58597	Human polynucleoti	
22	105	10.4	1808	23	AAH79353	DNA encoding novel	
23	102.5	10.2	1583	22	AAI60383	Human polynucleoti	
24	102.5	10.2	1909	22	ABA08598	Human Ser/Arg rich	
25	102	10.1	2695	24	AAH39557	Human SR-cyp DNA.	
26	102	10.1	10369	22	AAH46304	Tumour suppressor	
27	102	10.1	10369	24	ABL32393	Human immune syste	
c	28	101	10.0	886	24	ABQ35134	Oligonucleotide fo
29	101	10.0	886	24	ABQ35135	Oligonucleotide fo	
30	101	10.0	1527	21	AAH70121	Plasmodium faicipa	
31	101	10.0	2220	24	AAH24393	Human RNA metaboli	
c	32	100	9.9	1236	24	ABQ29236	Oligonucleotide fo
33	100	9.9	1236	24	ABQ29237	Oligonucleotide fo	
34	99.5	9.9	1454	24	ABR84288	Human cDNA differe	
35	99	9.8	1411	21	AAH75975	Human ORFX ORF1530	
36	99	9.8	2076	24	ABK83839	Human cDNA differe	
37	99	9.8	2736	24	ABK84078	Human cDNA differe	
38	99	9.8	3484	21	AAH78163	Human cancer assoc	
c	39	98.5	528	24	ABO50644	Oligonucleotide fo	
40	98.5	9.8	528	24	ABO50645	Oligonucleotide fo	
c	41	98.5	9.8	597	24	ABQ36486	Oligonucleotide fo
42	98.5	9.8	597	24	ABQ36487	Oligonucleotide fo	
43	98.5	9.8	1011	23	ABL24287	Drosophila melanog	
c	44	98.5	9.8	3011	23	ABL24286	Drosophila melanog
45	98	9.7	1536	21	AAH76106	Human ORFX ORF1661	

ALIGNMENTS

RESULT 1
AAA51368
ID AAA51368 standard; DNA; 579 BP.
XX
AC AAA51368;
XX
DT 26-SEP-2000 (first entry)
XX
DE Streptococcus agalactiae pho3-1 gene.
XX
KW Group B Streptococcus; pho3-1; virulence; vaccine; uropathic;
KW gene therapy; veterinary; mastitis; immunostimulant; antibacterial; ds.
XX
OS Streptococcus agalactiae.
XX
FH Key Location/Qualifiers
FT CDS 1..579
FT /*tag= a

THIS PAGE BLANK (USPTO)

Best Local Similarity: 22.04% Mismatches: 66
Query Match: 7.93% Indels: 58
DB: 4 Gaps: 7

US-09-868-352-23 (1-192) x US-09-134-001C-2695 (1-789)

```
Qy 4 GluValThrProGluMetLeuAsnTyrAsnLysTyrProGlyProGlnPheIleHisPhe 23
   |||:::  ::  ||| |||||  |||  |||
Db 307 GAAATTCACGACGATATTCCCAATGTTGGAAATAT-----AATTTTATACCTTT 357
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Qy 24 GluAsnIleValLysSerAspIleGluPheGlnLeuValIleAsnGluLysSerAla 43
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 358 GAAGTCTTCGAAAGTCTAAT-----AACGATCAACTACAT 393
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Qy 44 PheAspValThrValPheGlyGlnArgPheSerGluIleLeuLeuLysTyrAspPheIle 63
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 394 GTAACATCCACTCATTTTAAAGCAACCTATATTAGATAGTTACTTAAACACT----- 447
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Qy 64 ValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyr----- 78
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 448 -----AAAGAAACATTTATCTATACATGGATTCCTCGGAGATGACCTATT 492
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Qy 79 -----LysAspAlaSerThrIleArgLysAsnSerArgIleSerArgLeu 93
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 493 GTTTATATCGGAGGTAAGATAAAGAGATGTCCTCACTCTATCACGAAAGAAATTACGTAAA 552
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Qy 94 GluAspTyrIleLysGluTyrCysAsnPheGlyCysAlaTyrPheValLeuGluAsnPro 113
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 553 AAAGACTTTACT-----GTAAAGAAGTCCT 579
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Qy 114 AsnProArgAspIleLysPheAspGluArgProHisLysArgArgLysSerArgSer 133
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 580 AACAAATCGATCGCAAGTCATCAGATAACATACGTAATAA----- 621
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Qy 134 LysSerGlnSerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAla 153
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 622 -----AATGAATCTAACTCAGGTGTCCAAATTA 648
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Qy 154 HisPheThrSerLysArgLysAspThrLysArg-----ArgGlnGluArgHisIle 171
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 649 GAATCAGGCTGCATTGCGAAACACAGTTTTCAAACATTATAAATTAGATCGACATACA 708
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Qy 172 LysGluGluGlnAspLys 177
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 709 CGTAGTCATTCAGATAAA 726
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
```

Search completed: February 25, 2003, 12:53:56
Job time : 75 secs

Score: 81.50 Matches: 35
Percent Similarity: 39.71% Conservative: 19
Best Local Similarity: 25.74% Mismatches: 52
Query Match: 8.08% Indels: 31
DB: 3 Gaps: 4

US-09-868-352-23 (1-192) x US-08-680-897-1 (1-9880)

QY 67 TrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyrlYsAspAlaSerThr----- 83
Db 8138 TGGAAACAGCGCTCCAGAGATAGAGGGAGTCTCGGACACGCCAGCAAGCGGAAG 8197
QY 84 IleArgLysAsnSerArgIleSerArgLeuGluAspTyrlLysGluTyrcysAsnPhe 103
Db 8198 ATACGGAAGCAGAGGAGACAGAGG---AAGAGGCACCGAAGACATGCTGCAGTGCC 8254
QY 104 GlyCysAlaTyrcysPheValLeuGlu----- 111
Db 8255 GGGTGTGGCGGGAATATATCAGGCCATCAGCGGCCCATCTCCAGACCCCATGATATGA 8314
QY 112 -----AsnProAsnProArgAspIleLys 119
Db 8315 CGATTACAGGACTACGGCAGCAGTACGGTGAGGATGTCCCCCTTCGGCGCGGATGATAAA 8374
QY 120 PheAspAspGluArgProHisLysArgArgLysSerArgSerLysSerGlnSerLys 139
Db 8375 CATCAGCAACGAGTCCCGT---AAACGCAAGGTACGAGGTTGCGAGATCCGCTCTCAGCG 8432
QY 140 SerGlnThrArgAsnAsnArgSerGlnSerAsnAlaHisPheThrSerLysLys 159
Db 8433 GCGGACAAAGAAACAGGGCGGATACAGAGCGGTGATGATAGGAGATCGAGCGTCTCGAG 8492
QY 160 ArgLysAspThrLysArgArgGlnGlu-----ArgHisLysLysGlu 173
Db 8493 AGGCAGAACGACGCTCTGGCCAGGAGTCTCTACCCACGTCGAGCAG 8540

RESULT 37
US-08-961-527-167/c
; Sequence 167, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 167:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9711 base pairs

; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-167

Alignment Scores:

Pred. No.: 28.8 Length: 9711
Score: 81.00 Matches: 35
Percent Similarity: 43.36% Conservative: 27
Best Local Similarity: 24.48% Mismatches: 47
Query Match: 8.03% Indels: 34
DB: 4 Gaps: 8

US-09-868-352-23 (1-192) x US-08-961-527-167 (1-9711)

QY 60 TyrAspPheIleValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyrlYs 79
Db 1214 TTTAACTATCTTGGCTTGGACTATGGACTCAAAACAAGTCGCAATCTCAGGACTTTCTCCA 1155
QY 80 AspAlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrlLysGlu 99
Db 1154 GATCCAGAGCCA-----TCAGCTGCTGCTTGCAGATTTCAGAGATATGTCAAGAAA 1101
QY 100 TyrCysAsnPheGlyCysAlaTyrcysPheValLeuGluAsnProAsnProArgAspIleLys 119
Db 1100 AAT---AAAATCGCCTATATCTATTT----- 1077
QY 120 PheAspAspGluArgProHisLys-----ArgArgLysSerArgSer 133
Db 1076 -----GAAGAAATCGCTCAACAGCCCTTGTAAACACTTTCAAAAGACGAGCGTGTCA 1023
QY 134 LysSerGlnSerLysSer-----GlnThrArgAsnAsnArgSerGlnSer 149
Db 1022 AAACGTGATGCTCAATCCTTTAGAAAGCTGCACAGAGAGCACCAGGCTGGAGAAA 963
QY 150 AsnAlaAsnAla-----HisPheThrSerLys-----LysArgLysAspThrLysArg 165
Db 962 ACTACATTTTCGTGATGGAGAAAAACCTCAAGGCTTTGAAACAAACAACACAGCAAGAAG 903
QY 166 ArgGlnGluArgHisLysLysGluGlu-----GlnAspLysGluMetThrSer 181
Db 902 GCCCAAGAAATGAACCTGAAAGGAGGAGGATACCAAGACAGTCCAAAATGGTTACTTCG 843
QY 182 AlalysGln 184
Db 842 AGGATGCAG 834

RESULT 38

US-08-691-563C-46
; Sequence 46, Application US/08691563C
; Patent No. 6001987

; GENERAL INFORMATION:
; APPLICANT: Herve PERRON
; APPLICANT: Frederic BESEME
; APPLICANT: Frederic BEDIN
; APPLICANT: Glauca PARANHOS-BACCALA
; APPLICANT: Florence KOMURIAN-PRADEL
; APPLICANT: Colette JOLIVET
; APPLICANT: Bernard MANDRAND
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
; TITLE OF INVENTION: THERAPEUTIC PURPOSES
; NUMBER OF SEQUENCES: 92
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oliff & Berridge
; STREET: 700 South Washington Street, Suite 300
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible


```
; LOCATION: 13149..14171
US-09-299-268-14

Alignment Scores:
Pred. No.: 32.6 Length: 14176
Score: 82.50 Matches: 35
Percent Similarity: 41.67% Conservative: 25
Best Local Similarity: 24.31% Mismatches: 45
Query Match: 8.18% Indels: 39
DB: 4 Gaps: 8

US-09-868-352-23 (1-192) x US-09-299-268-14 (1-14176)
QY 5 ValThrProGluMetLeuAsn-----TyrAsnLysTyrProGlyProGln 19
Db 9796 GTTGACCTAATCATAGTAAGTGGCATCTATTATTATAGCGAATAC-----GAA 9846
QY 20 PheIleHisPheGluAsnIleValLysSerAspAspIleGluPheGlnLeuVal----- 37
Db 9847 TATCTTTTAAAAAATCAAAATTATAAAATAAGACATAAATATATATATAGATAAG 9906
QY 38 IleAsnGluLysSerAlaPheAspValThrValPhe----- 49
Db 9907 ATTAATAATATACATAGTATAGATATATATATATATATATATATATATATATATAT 9966
QY 50 -----GlyGlnArgPheSerGluIleLeuLysTyrAsp--- 61
Db 9967 TATAACAACACATGCCCATGTGTAGTATTATTAAGGAA---CTCATGAATATGATATA 10023
QY 62 -----PheIleValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPhe 77
Db 10024 TTAGTACAAATATATATATATATATATATATATATATATATATATATATATAT 10062
QY 78 TyrLysAspAlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIle 97
Db 10063 TACAAAGAGGGAGATACTATATCCATTACATACAGTTTAAATGTAAGATGATATAATT 10122
QY 98 LysGluTyrCysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAsp 117
Db 10123 AAA---CATTGTAAGTCTCTTAGTATGTTTGTCTATATATATATATATATATATAT 10179
QY 118 IleLysPheAsp 121
Db 10180 GTAGATTTTGAT 10191

RESULT 32
US-09-086-912-1
; Sequence 1, Application US/09086912
; Patent No. 6166192
; GENERAL INFORMATION:
; APPLICANT: Bruce M. Spiegelman, Pere Puigserver and Zhidan Wu
; TITLE OF INVENTION: PGC-1, A No. 6166192el Brown Pat PPAR(SYMBOL
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESSES:
; ADDRESS: LAHIVE & COCKFIELD, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/086,912
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/048,107
; FILING DATE: 30-MAY-1997
; ATTORNEY/AGENT INFORMATION:
```

```
; NAME: Mandragouras, Amy B.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: DFN-023
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3066 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 92..2482
US-09-086-912-1

Alignment Scores:
Pred. No.: 4.15 Length: 3066
Score: 82.00 Matches: 38
Percent Similarity: 37.57% Conservative: 30
Best Local Similarity: 20.99% Mismatches: 49
Query Match: 8.13% Indels: 64
DB: 4 Gaps: 8

US-09-868-352-23 (1-192) x US-09-086-912-1 (1-3066)
QY 65 GlyAspTrpGlyAsnGluGln-----LeuArg 73
Db 1547 GCGGAGCTTCAGTAATGAACAATCTCCAACTACCTGTGTTTATAAATTCAGGACTAGCC 1606
QY 74 LeuArgGlyPheTyrLysAspAlaSerThrIleArgLysAsnSerArgIleSerArgLeu 93
Db 1607 ATGGATGCGCTATTTCATGCACAGTGAAGAT-----GAAAGTGATAAACTGACCTACCCT 1660
QY 94 GluAspTyrIleLysGluTyrCysAsnPhe-----GlyCysAlaTyrPhe--- 108
Db 1661 TGGATGGCAGCAGCAGCCCTATTTCATTTTCGATGTCTGCTCTTCTTCTTCTTAAAC 1720
QY 109 -----ValLeuGluAsnProAsnProArgAspIleLysPheAspAspGluArgProHis 126
Db 1721 TCTCCGTGTCGAGACACAGTGTGTCACCCGAAATCTTATTT---TCTCAAGACCCCAA 1777
QY 127 LysArgArg----- 129
Db 1778 AGGATGCGCTCTCTTCAAGATCCTTTTCTCGACACAGGTGCGTGTCCCGATCACCATAT 1837
QY 130 ---LysSerArgSerLysSerGlnSerLysSerGlnThrArgAsn----- 144
Db 1838 TCCAGGTCAAGATCAAGGTCCCGAGCAGTAGATCCTCTTCAAGATCTGTACTACTAT 1897
QY 145 -----AsnArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLys 158
Db 1898 GAATCAAGCCACTACAGACACCCGACACCCGCAATCTCCCTTGTGTGTGAGATCAGT 1957
QY 159 LysArgLysAspThrLysArg----- 166
Db 1958 TCAAGGTCACTACAGCCGTAGGCCCGCCAGGTACGACAGCTATGAAGCCTATGAGCAGAA 2017
QY 167 -----GlnGluArgHisIleLysGluGluAsnLysGluMetThrSerAlaLys 193
Db 2018 AGGCTCAAGAGGGATGAATACCGCAAAAGACGAGAGCGGAGGTCTCGAAAGGGCCAA 2077
QY 184 Gln 184
Db 2078 CAG 2080

RESULT 33
US-09-203-453-1
; Sequence 1, Application US/09203453
; Patent No. 6426411
; GENERAL INFORMATION:
```



```

Db 817 TCTACATCTAAAGACACCAAGTTAGAAAAGAAACTACAGTCAAGAAAGGAATCT 876
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 176 AspLysGluMetThrSerAlaLysGln 184
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 877 AAAGCCACACGACTAAAGTGAAGAA 903
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 27
US-09-402-328-1/c
; Sequence 1, Application US/09402328
; Patent No. 6365728
; GENERAL INFORMATION:
; APPLICANT: Purdue Research Foundation,
;           Hodges, Thomas K.
;           Lysnik, Leszek A.
; TITLE OF INVENTION: Regulatory Element For Expressing Genes
;           In Plants
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Barnes & Thornburg
; STREET: 11 S. Meridian
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/09/402,328
; FILING DATE: 05-No. 6365728-1999
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Breen, John P.
; REGISTRATION NUMBER: 38,833
; REFERENCE/DOCKET NUMBER: 3220-29933
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 231-7745
; TELEFAX: (317) 231-7433
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5285 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Arabidopsis thaliana
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-402-328-1
Alignment Scores:
Pred. No.: 7.89 Length: 5285
Score: 82.50 Matches: 37
Percent Similarity: 39.33% Conservative: 22
Best Local Similarity: 24.67% Mismatches: 52
Query Match: 8.18% Indels: 39
DB: 4 Gaps: 6

US-09-868-352-23 (1-192) x US-09-402-328-1 (1-5285)
QY 34 PheGlnLeuValIleAsnGluLysSerAlaPheAspValThrValPheGlyGlnArgPhe 53
    :||| ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2372 TATCTTTATTTTGTCCAGTATAAAAAGCTCGATGCACAGAGTTTCGGAATCGGTTTT 2313
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 54 SerGluIleLeuLysTyrAspPheIleValGlyAspTrpGlyAsnGluGlnLeuArg 73
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2312 -----CTCATCAAT 2304
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```

QY 74 LeuArgGlyPheTyrLysAspAlaSerThrIleArgLysAsnSerArgIleSerArgLeu 93
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2303 ATGTCGGGTATGTTAAAGAC-----AAGAATTCT----- 2274
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 94 GluAspTyrIleLysGluTyrCys-----AsnPheGlyCysAlaTyrPheValLeuGlu 111
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2273 GAGAGCTGCATTCGTTTATTGCACTTAACATTTCCCTATGTCATTACAGGTTAGAGAA 2214
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 112 AsnProAsnProArgAspIleLys-----PheAspAspGluArg 124
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2213 TATGATTCAAGGAGGATTCGAGGAGCCCGCGTGGAGATCCTATTCTAGAGCCGC 2154
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 125 ProHisLysArgLysSerArgSerLysSerGlnSer-----SerLysSerGln 141
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2153 AGCCGTGGACGAGTCCCGCTAGTCGTAGCCGAGCAAGAGCAGGAGCAAGAGCAGA 2094
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 142 ThrArgAsnAsnArgSerGlnSerAsnAlaHisPheThrSerLysLysArgLys 161
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2093 AGTCCAAAGGCTAAATCTTTGCGTAGATCGCGTCAAAATCTACATCGAGATCTCCTCGC 2034
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 162 AspThrLysArgArgGlnGluArgHisIle 171
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2033 TCTCGCTCCCGCTCTAAGTCGAGGTCACTG 2004
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 28
US-08-307-499-1/c
; Sequence 1, Application US/08307499
; Patent No. 5651972
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Vi uela, Eladio
; APPLICANT: Gibbs, E.P.J.
; TITLE OF INVENTION: Use of Recombinant Swine Poxvirus as a
;           Live Vaccine Vector
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: U.S.A.
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/307,499
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/908,241
; FILING DATE: 1-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/908,630
; FILING DATE: 29-JUN-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/342,212
; FILING DATE: 21-APR-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UF35.1.FWCC1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14176 base pairs
; TYPE: nucleic acid
```

```
Alignment Scores:
Pred. No.: 3.74 Length: 3143
Score: 82.50 Matches: 42
Percent Similarity: 36.51% Conservative: 27
Best Local Similarity: 22.22% Mismatches: 47
Query Match: 8.18% Indels: 73
DB: 2 Gaps: 9

US-09-868-352-23 (1-192) x US-08-973-831-1 (1-3143)
Qy 14 LysTyrProGlyProGlnPheIleHisPheGluAsnIleVal-----LysSerAsp 30
Db 502 CGTTATTAAATCCATTATTTGCTTCATATAGCATATTCCTATGAGTTCATCAGAC 561
Qy 31 AspileGluPheGlnLeuValIleAsnGluLysSerAlaPheAspValThrValPheGly 50
Db 562 GAAGAAGACATTGCCTTCCTAGACTCCTATAAATCATCTCGATCACT----- 612
Qy 51 GlnArgPheSerGluIleLeuLeuLysTyrAspPheIleValGlyAspTrpGlyAsnGlu 70
Db 612 ----- 612
Qy 71 GlnLeuArgLeuArgGlyPheTyrLysAspAlaSerThr----- 83
Db 613 -----TCAGCTTCCACTTATGAAGACGATGAAGATGAT 645
Qy 84 -----IleArgLysAsnSerArgIleSerArgLeuGlu---AspTyrIleLysGlu 99
Db 646 GATATCCCTTTAGCTAAATAATCCAGGAAAGAGGGTGAATCTGATTATGAAGAAGAT 705
Qy 100 TyrCysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspileLys 119
Db 705 ----- 705
Qy 120 PheAspAspGluArgProHisLysArgArgLys-----SerArgSerLysSerGln 136
Db 706 ---GAACACGAAGTCCCATTTGAAAGAGAGAAATTTGTCTAATGCTAGACGAAACAA 762
Qy 137 SerSerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPheThr 156
Db 763 ---GTTAAACCGAACTAAAGTTAAAGAGAACTAAAGTAAATATCC---AAA 816
Qy 176 AspileGluMetThrSerAlaLysGln 184
Db 877 AAAGCCACAGCACATAAGTGAAGAA 903

RESULT 26
PCT-US96-09530A-1
; Sequence 1, Application PC/TUS9609530A
; GENERAL INFORMATION:
; APPLICANT: Kmiec, Eric B.
; APPLICANT: Gerhold, David L.
; APPLICANT: Strauss, Allyson Cole
; TITLE OF INVENTION: Anti-fungal Agents and Methods of
; TITLE OF INVENTION: Identifying and Using the Same
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESS: Woodcock Washburn Kurtz Mackiewicz & Norris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 6.1
; CURRENT APPLICATION DATA:
```

```
US-08-485-621-1
; Sequence 1, Application US/08485621
; Patent No. 5691187
; GENERAL INFORMATION:
; APPLICANT: Kmiec, Eric B.
; APPLICANT: Gerhold, David L.
; APPLICANT: Strauss, Allyson Cole
; TITLE OF INVENTION: Anti-fungal Agents and Methods of
; TITLE OF INVENTION: Identifying and Using the Same
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
; ADDRESSEE: No. 5691187ris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,621
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TJU-1584
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3143 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: genomic DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 547..2889
;
US-08-485-621-1
Alignment Scores:
Pred.No.: 3.74 Length: 3143
Score: 82.50 Matches: 42
Percent Similarity: 36.51% Conservative: 27
Best Local Similarity: 22.22% Mismatches: 47
Query Match: 8.18% Indels: 73
DB: 1 Gaps: 9
US-09-868-352-23 (1-192) x US-08-485-621-1 (1-3143)
Qy 14 LystyrProGlyProGlnPheIleHisPheGluAsnIleVal-----
::||| ||| |||::: ::|||:::
Db 502 CGTATTAAATCCATTATTGTTCGTATAGCATAAATTCCTATGCA
Qy 31 AspleGluPheGlnLeuValIleAsnGluLysSerAlaPheAspVal
::: ::: ||| ::| ::|
Db 562 GAAGAAGACATTGCCTTGCTAGACTCGCTAAAAAATCATCCTCGTGCATCA
Qy 51 GlnArgPheSerGluIleLeuLeuLysTyrrAspPheIleValGlyAspThr
Db 612 -----
Qy 71 GlnLeuArgLeuArgGlyPheTyrrLysaspAlaSerThr-----
Db 613 -----TCAGCTTCCACTTAATGAAGACG
Qy 84 -----IleArqLYsAsnSerArgIleSerArqLeuGlu---AspTr
```

```

Db      646  GATATCCCTTTAGTCAAAAATCCAGGAAAGAGGGTTGAAATCTCATATGAAGAAGAT 705
Qy      100  TyTcysAsnPheGlyCysAlaTyRpheValLeuGluAsnProAsnProArgAspIleLys 119
Db      705  -----
Qy      120  PheAspAspGluArgProHisLysArgArgLys-----SerArgSerLysSerGln 136
Db      706  --GAACAGAGTCCCATTTGAAAAAGAGAAAAATTGCTTAATGGTAGACCAAAAAACAA 762
Qy      137  SerSerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPheThr 156
Db      763  ---GTTAAACCCGAACTAAAGTTAAAGAGGAACCTTAAAGTGCCCAATAATCC---AAA 816
Qy      157  SerLysLysArgLysAspThrLysArgArgGlnGluArg---HisIleLysGluGln 175
Db      817  TCTACATCTAAAAAGGACACCAAAAGTTAAGAAAGAGAAACTACAGTCAAGAAGGAATCT 876
Qy      176  AspLysGluMetThrSerAlaLysGln 184
Db      877  AAAGCCACAGCACCTAAAGTCAAGAA 903

RESULT 25
US-08-973-831-1
; Sequence 1, Application US/08973831
; Patent No. 5942386
; GENERAL INFORMATION:
; APPLICANT: Kmiec, Eric B.
; APPLICANT: Gerhold, David L.
; APPLICANT: Strauss, Allyson Cole
; TITLE OF INVENTION: Anti-fungal Agents and Methods of
; TITLE OF INVENTION: Identifying and Using the Same
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
; ADDRESSEE: No. 5942386ris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/973,831
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/485,621
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: DeLuca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TJU-1584
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3143 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: genomic DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 547..2889
; US-08-973-831-1

```

US-08-973-831-1

[illegible]


```
Qy 133 rLysSerGlnSerLysSerGlnThrArgAsnAsn-----ArgSerGlnSerAs 150
||| :|||:|||||:|||||: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 747 AGGCTGAGAGAGTACGAGACAGATACAGGAAAAATTCAGCGAGGAGGAGAAAGAGAA 806
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 150 nAlaAsnAlaHisPheThrSerLysLysArgLysAspThrLysArgGlnGluArgHi 170
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 807 C-----GGCAAGAACAGGAAAGACACAAAAGGAAACGAGAAAGAAATG 848
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 170 sileLysGluGlnAspLysGluMetThrSerAlaLys 183
|||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 849 GAACAAGAAAGGATAAGACAAGATATGATGAAAAGGAAA 888
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
RESULT 19
US-08-460-428A-1
; Sequence 1, Application US/08460428A
; Patent No. 5912337
; GENERAL INFORMATION:
; APPLICANT: Tripp, Cynthia A.
; APPLICANT: Frank, Glenn R.
; APPLICANT: Grieve, Robert B.
; TITLE OF INVENTION: NOVEL PARASITIC HELMINTH
; TITLE OF INVENTION: P22U PROTEINS
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln St., Suite 3500
; CITY: Denver
; STATE: CO
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,428A
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-13-3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/863-9700
; TELEFAX: 303/863-0223
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 913 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3..911
US-08-460-428A-1
Alignment Scores:
Pred. No.: 0.551 Length: 913
Score: 83.00 Matches: 41
Percent Similarity: 42.21% Conservative: 24
Best Local Similarity: 26.62% Mismatches: 42
Query Match: 8.23% Indels: 47
DB: 2 Gaps: 9
US-09-868-352-23 (1-192) x US-08-460-428A-1 (1-913)
Qy 68 GlyAsnGluGlnLeuArgLeuArgGlyPheThrLysAspAlaSerThrIleArg---Lys 86
||||||| :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 456 GGTAACGAATCAATCAATATATCAGA-----TCTGATCCGGATCAATTTTCGATGTGAA 506
||||||| :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Qy 87 AsnSerArg---IleSerArgLeuGluAspTyrIleLysGluTyr---CysAsnPheGly 104
||| :|||:|||||:|||||: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 507 AATGCAAAATGTATCGCACAAATTCATCGATGTAATCAAAATATGATTTGTGATGATGGT 566
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 105 -----CysAlaTyrPheValLeu----- 110
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 567 ACAGATGAACAACTTGTGTAATATTTCTGCAAGCTTTTGCACAAAGCGAGAGGTGTAACG 626
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 111 -----GluAsnProAsnProArg----- 116
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 627 GTGAGGATAATGCAATTCGAGATGACGAGATACCAATATTACTGTATCCATGCAACAG 686
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 117 -----AspIleLysPheAspGluArg-ProHisLysArgArgLysSerArgSe 133
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 687 AATACGATCAAGTAAGGAAGATAAGGCGCGAATGCAAGAGGAGGAGGAGGAGAA 746
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 133 rLysSerGlnSerLysSerGlnThrArgAsnAsn-----ArgSerGlnSerAs 150
|||: |||:|||||:|||||: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 747 AGGCTGAGAGAGTACGAGGAAACAGATACAGCAAAATTCAGCGAGGAGGAGAAAGAGAA 806
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 150 nAlaAsnAlaHisPheThrSerLysLysArgLysAspThrLysArgArgGlnGluArgHi 170
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 807 C-----GGCAAGAACAGGAAAGACACAAAAGGAAACGAGAAAGAAATG 848
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 170 sileLysGluGlnAspLysGluMetThrSerAlaLys 183
|||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 849 GAACAAGAAAGGATAAGACAAGATATGATGAAAAGGAAA 888
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
RESULT 20
US-08-458-860A-1
; Sequence 1, Application US/08458860A
; Patent No. 6100390
; GENERAL INFORMATION:
; APPLICANT: Frank, Glenn R.
; APPLICANT: Tripp, Cynthia A.
; APPLICANT: Grieve, Robert B.
; TITLE OF INVENTION: NOVEL PARASITIC HELMINTH
; TITLE OF INVENTION: P22U NUCLEIC ACID MOLECULES
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln St., Suite 3500
; CITY: Denver
; STATE: CO
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,860A
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-13-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/863-9700
; TELEFAX: 303/863-0223
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 913 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3..911
US-08-458-860A-1
```



```

1  NUMBER OF SEQUENCES: 63
2  CORRESPONDENCE ADDRESS:
3  ADDRESSEE: Lowe, Price, LeBlanc & Becker
4  STREET: Suite 300, 99 Canal Center Plaza
5  CITY: Alexandria
6  STATE: Virginia
7  COUNTRY: USA
8  ZIP: 22314
9
10 COMPUTER MODEL:
11 MEDIUM TYPE: Floppy disk
12 COMPUTER: IBM PC compatible
13 OPERATING SYSTEM: PC-DOS/MS-DOS
14 SOFTWARE: DOS Text File
15
16 CURRENT APPLICATION DATA:
17 APPLICATION NUMBER: US/07/799,828C
18 FILING DATE: 19911127
19 CLASSIFICATION: 435
20 ATTORNEY/AGENT INFORMATION:
21 NAME: D.J. Mills
22 REGISTRATION NUMBER: 34,506
23 REFERENCE/DOCKET NUMBER: 717081A
24 TELECOMMUNICATION INFORMATION:
25 TELEPHONE: 703 684 1111
26 INFORMATION FOR SEQ ID NO: 7:
27 SEQUENCE CHARACTERISTICS:
28 LENGTH: 291
29 TYPE: NUCLEIC ACID
30 STRANDEDNESS: single
31 TOPOLOGY: linear
32 MOLECULE TYPE: DNA (genomic)
33
34 US-07-799-828C-7

```

Alignment Scores:	
Pred. No.:	0.046
Score:	86.00
Percent Similarity:	36.70%
Best Local Similarity:	22.02%
Query Match:	8.52%
DB:	1
Gaps:	2
Indels:	3
Mismatches:	37
Conservative:	16
Matches:	24
Length:	291

US-09-868-352-23 (1-192) x US-07-799-828C-7 (1-291)

Qy	79	LysAspAlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLys	98
		: : : : :	
Db	43	AAAGAGACAAGAGAGTTAGACAAAGAAAGA-	75
Qy	99	GluTyrCysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspIle	118
Db	76	-----GAGAGAGAGAGAAGGAAGGAAGG	99
Qy	119	LysPheAspAspGluArgProHisLysArgArgLysSerArgSerLysSerGlnSerSer	138
		: : : : : : : : : : : :	
Db	100	AGCAAAAGAAAGAAAGAAAGAAAGAGAGAAAGAAAGAAAGAAAGAAAGA	159
Qy	139	LysSerGlnThrArgAsnAsnArgSerGlnSerAsnAlaHisPheThrSerLys	158
		: : : :	
Db	160	AGCAAAAGAAAGAAAGAAAGA-	186
Qy	159	LysArgLysAspThrLysArgArgGlnGluArgHisIleLysGluGlnAspLysGlu	178
		: : : : : : : : : : : : : : : : : : : :	
Db	187	AAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGA	246
Qy	179	MetThrSerAlaLysGlnHisLeuLeu	187
Db	247	AGCAAGGAAAGAAAGAGCAAGTTACTA	273

RESULT 10

US-08-074-275-7
; Sequence 7, Application US/08074275
; Patent No. 5468610
; GENERAL INFORMATION:
; APPLICANT: Drs. Carl R. Merrill and
; APPLICANT: Mihal H. Polymeropoulos

```

; TITLE OF INVENTION: THREE HIGHLY INFORMATIVE REPEAT
; TITLE OF INVENTION: POLYMORPHIC DNA MARKERS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lowe, Price, LeBlanc & Becker
; STREET: Suite 300, 99 Canal Center Plaza
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: DOS Text File
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/074,275
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/707,501
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: J.G. Mullins
; REGISTRATION NUMBER: 33073
; REFERENCE/DOCKET NUMBER: 717081
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703 684 1111
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 291
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-074-275-7

```

Alignment Scores:		
Pred. No.:	0.046	291
Score:	86.00	24
Percent Similarity:	36.70%	16
Best Local Similarity:	22.03%	37
Query Match:	8.52%	32
DR:	1	2
	Length:	
	Matches:	
	Conservative:	
	Mismatches:	
	Indels:	
	Gaps:	

US-09-868-352-23 (1-192) X US-08-074-275-7 (1-291)

Qy	79	LysAspAlaSerThrIleArgLysAsnSerArgGileSerArgLeuGluAspTyrIleLys	98
Db	43	AAAGAGACAAAGAGATTAGAAAAGAAGA-	75
Qy	99	GluTyrCysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspIle	118
Db	76	- - - - - - - - - - GAGAGAGAGAAGGAAGGAAGG	99
Qy	119	LysPheAspGluArgProHisLysArgGlcLysSerArgSerLysSerGlnSerSer	138
Db	100	AAGAAAAAAGAAAGAAAGAAAGAGAGAGAGAGAAAGAAAGAAAGAAAGA	159
Qy	139	LysSerGlnThrArgAsnAsnArgSerGlnSerAsnAlaHisPheThrSerLys	158
Db	160	AGCAAGAACAAAGAAAGAAACA- - - - - AAG	186
Qy	159	LysArgLysAspThrLysArgArgGlnGluArgHisIleLysGluGlnAspLysGlu	178
Db	187	AAAAAGAAACAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA	246
Qy	179	MetThrSerAlaLysGlnHisLeuLeu	187
Db	247	AGGAAGCAAGAAAGAGCAAGTTACTA	273

RESULT 11

US-08-480-366-7

```
; STREET: Suite 300, 99 Canal Center Plaza
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: USA
; Zip: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: DOS Text File
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/922,723A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: D.J. Mills
; REGISTRATION NUMBER: 34506
; REFERENCE/DOCKET NUMBER: 717081B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703 684 1111
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 291
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-07-922-723A-7

Alignment Scores:
Pred. No.: 0.046 Length: 291
Score: 86.00 Matches: 24
Percent Similarity: 36.70% Conservative: 16
Best Local Similarity: 22.02% Mismatches: 37
Query Match: 8.52% Indels: 32
DB: 1 Gaps: 2

US-09-868-352-23 (1-192) x US-07-922-723A-7 (1-291)
QY 79 LysAspAlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLys 98
Db 43 AAAGACAAAGACAGAGATTAGAAGAAAAGAAAGAA----- 75
QY 99 GluTyrCysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspIle 118
Db 76 -----GAGAGACAGACAAGGAAGGAGCG 99
QY 119 LysPheAspAspGluArgProHisLysArgArgLysSerArgSerLysSerGlnSerSer 138
Db 100 AAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAA 159
QY 139 LysSerGlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLys 158
Db 160 AAGAAAGAAAGAAAAAGAAAAAGAAAAAGAA----- 186
QY 159 LysArgLysAspThrLysArgArgGlnGluArgHisIleLysGluGluGlnAspLysGlu 178
Db 187 AAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 246
QY 179 MetThrSerAlaLysGlnHisLeuLeu 187
Db 247 AGGAAGGAAAGAAAGCAAGTACTA 273

RESULT 9
US-07-799-828C-7
; Sequence 7, Application US/07799828C
; Patent No. 5378602
; GENERAL INFORMATION:
; APPLICANT: Drs. Carl R. Merrill and
; APPLICANT: Michael H. Polymeropoulos
; TITLE OF INVENTION: TWENTY SEVEN HIGHLY INFORMATIVE
; TITLE OF INVENTION: MICROSATELLITE REPEAT
; TITLE OF INVENTION: POLYMORPHIC DNA MARKERS
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lowe, Price, LeBlanc & Becker
```


DB: 2 Gaps: 2

US-09-868-352-23 (1-192) x US-08-718-661-1 (1-2790)

QY 130 LysSerArgSerLysSerGlnSer---SerLysSerGlnThrArgAsn---AsnArgSer 147
 Db 2295 ::::||||:::||||||||| ||:::||||:::||||| ||||||| 2355
 QY 148 GlnSerAsnAlaAsnAlaHisPheThrSerLysArgLysAspThrLysArgArgGln 167
 Db 2355 :||||| ::: ||| ||||| ||||| ||||| 2414
 QY 168 GlnArgHisLysGluGlnAspLysGluMetThrSerAlaLysGlnHisLeuLeu 187
 Db 2415 ::: ::::|||||::: ||||| 2474
 QY 188 PheValArgLysAsn 192
 Db 2475 AGGAAGCGGAAGAGC 2489

RESULT 5

US-09-018-584A-27

; Sequence 27, Application US/09018584A

; Patent No. 6238863

; GENERAL INFORMATION:

; APPLICANT: Schumm, James W.

; APPLICANT: Bachter, Jeffery W.

; TITLE OF INVENTION: MATERIALS AND METHODS FOR

; TITLE OF INVENTION: IDENTIFYING AND ANALYZING INTERMEDIATE TANDEM

; TITLE OF INVENTION: REPEAT DNA MARKERS

; NUMBER OF SEQUENCES: 147

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Promega Corporation

; STREET: 2800 Woods Hollow Road

; CITY: Madison

; STATE: Wisconsin

; COUNTRY: U.S.A.

; ZIP: 53711-5399

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette - 3.5 inch, 1.44 Mb

; COMPUTER: IBM compatible PC

; OPERATING SYSTEM: Windows 95

; SOFTWARE: Word 97 (DOS text format)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/018,584A

; FILING DATE: 04-Feb-1998

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Grady J. Frenchick

; REGISTRATION NUMBER: 29,018

; REFERENCE/DOCKET NUMBER: 16026.9180

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (608) 257-3501

; TELEFAX: (608) 257-2275

; INFORMATION FOR SEQ ID NO: 27:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 333 bp

; TYPE: Nucleic Acid

; STRANDEDNESS: Double

; TOPOLOGY: Circular

; MOLECULE TYPE: Genomic DNA

; HYPOTHETICAL: no

; IMMEDIATE SOURCE:

; LIBRARY: plasmid, pGem3zf(+)

; CLONE: G539

; POSITION IN GENOME:

; CHROMOSOME/SEGMENT: 15q26.2

US-09-018-584A-27

Alignment Scores:

Pred. No.: 0.0367 Length: 333

Score: 87.50 Matches: 17

Percent Similarity: 52.86% Conservative: 20

;
; LENGTH: 5558 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-103

Alignment Scores:
Pred. No.: 5,72e-43 Length: 5558
Score: 438.00 Matches: 88
Percent Similarity: 68.8% Conservative: 36
Best Local Similarity: 48.8% Mismatches: 50
Query Match: 43.41% Indels: 6
DB: 4 Gaps: 4

US-09-868-352-23 (1-192) x US-08-961-527-103 (1-5558)

QY 1 MetArgLysGluValThrProGluMetLeuAsnTyrAsnLysTyrProGlyProGlnPhe 20
|||||
Db 4885 ATCGGAAAGAAATTCACCTGAATATACACTATAACAAGTTTCCTCGTCGGAGTTC 4826
QY 21 IleHisPheGluAsnIleValLysSerAspPheGluPheGlnLeuValIleAsnGlu 40
::: |||
Db 4825 CATTTACACGGGACACAGCTCGAAACGGAAGGATAGCTTTTCTTGTGTTGAAATATC 4766
QY 41 LysSerAlaPheAspValThrValPheGlyGlnArgPheSerGluIleLeuLysTyr 60
||| |||
Db 4765 AAGGATGCTCTTGCATGTGACGACTTTTAATACGCGTTTTCAGAGATATTAACCAAGTTT 4706
QY 61 AspPheIleValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAsp 80
|||||
Db 4705 GATTATATCTGGGGACTTGGAGCAACGACGCTTCGCTACGAGGTTTTCACAGGAT 4646
QY 81 AlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyr 100
||| |||
Db 4645 GACCGAACAAGAAATCTGAAATAATCAGTCGTTTACAGACTACCTTTTAGAGTAT 4586
QY 101 CysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspPheLysPhe 120
|||||
Db 4585 TGTAGTTATGGTTGTGCTCTATTTGCTTAGAAAATGAAGCCCTTAAGCGACATCATTT 4526
QY 121 AspAspGluArgProHisLysArgLysSerArgLysSerArgLysSerLysSer 140
||| |||
Db 4525 GAC-----AAGAAATGCGTAGACGGAGAGAAACACCTTCTAGAAAAGAAAGAA 4472
QY 141 GlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysLysArg 160
||| |||
Db 4471 CGGACTCAAACTAAACGA---AAGTCGAATGCAGATAAGAAA---AATAGACGTCGTCAG 4418
QY 161 LysAspThrLysArgArgGlnGluArgHisIleLysGluGlu-----GlnAspLysGlu 178
||||| |||
Db 4417 AAAGACGACAAGGGACAACGTCATTTTGTGTCATTCGTCAGAAAGTGTAGAGATAAGGAG 4358

RESULT 2
US-09-134-001C-2057
; Sequence 2057, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2057
; LENGTH: 399
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2057

Alignment Scores:
Pred. No.: 2.5e-22 Length: 399
Score: 254.50 Matches: 47
Percent Similarity: 68.93% Conservative: 24
Best Local Similarity: 45.63% Mismatches: 29
Query Match: 25.22% Indels: 3
DB: 4 Gaps: 2

US-09-868-352-23 (1-192) x US-09-134-001C-2057 (1-399)

QY 24 GluAsnIleValLysSerAspPheGluPheGlnLeuValIleAsnGluLysSerAla 43
|||||
Db 4 GAAATATGATTAAAGTCGACCAACAATATTTGAATTGATAGAGAATATAGAGAATGT 63
QY 44 PheAspValThrValPheGlyGlnArgPheSerGluIleLeuLysTyrAspPheIle 63
|||||
Db 64 TTGTATGAGGAATAATTTTCAGCTAGTATTCGATATATTACAGAAATATGATTATGTC 123
QY 64 ValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAspAlaSerThr 83
|||||
Db 124 GTAGGTGACTATGTTTACGATCAATTTACGCTTAAAGAGATTTTATAAGATAGTAATAA 183
QY 84 IleArgLys---AsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyrCysAsn 102
::: ||| |||
Db 184 AAGCGAGAATAAGTAACAGGATTTTCAAGTATATACAGATATATACAGAAATATGTAAT 243
QY 103 PheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspPhe-----LysPhe 120
|||||
Db 244 TTGGTTGTCTTATTTTGTAGTCAGACGATGTCACCAAAATGAATTTATTGAGAAATA 303
QY 121 AspAspGlu 123
|||||
Db 304 GATGATAAA 312

RESULT 3
US-08-931-999-4
; Sequence 4, Application US/08931999
; Patent No. 6043219
; GENERAL INFORMATION:
; APPLICANT: Iandolo, John J.
; APPLICANT: Crupper, Scott S.
; NUMBER OF INVENTION: Broad Spectrum Chemotherapeutic Peptide
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Hovey, Williams, Timmons & Collins
; STREET: 2405 Grand Boulevard, Suite 400
; CITY: Kansas City
; STATE: Missouri
; COUNTRY: U.S.A.
; ZIP: 64108
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/931,999
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/710,561
; FILING DATE: 19-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Collins, John M.
; REGISTRATION NUMBER: 26,262
; REFERENCE/DOCKET NUMBER: 25043-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 816/474-9050
; TELEFAX: 816/474-9057
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6755 base pairs

Result No.	Score	Query 8			DB	ID	Description
		Match	Length	Length			
C	1	438	43.4	5558	4	US-08-961-527-103	Sequence 103, Appl
	2	254.5	25.2	399	4	US-09-134-001C-2057	Sequence 2057, Ap
	3	95	9.4	6755	3	US-08-931-999-4	Sequence 4, Appli
	4	94	9.3	2790	2	US-08-718-661-1	Sequence 1, Appli
	5	87.5	8.7	333	4	US-09-018-584A-27	Sequence 27, Appl
	6	87	8.6	2017	1	US-07-667-276A-3	Sequence 3, Appli
	7	87	8.6	2237	4	US-08-914-999-7	Sequence 7, Appli
	8	86	8.5	291	1	US-07-922-723A-7	Sequence 7, Appli
	9	86	8.5	291	1	US-07-799-828C-7	Sequence 7, Appli
	10	86	8.5	291	1	US-08-074-275-7	Sequence 7, Appli
	11	86	8.5	291	1	US-08-480-366-7	Sequence 7, Appli
	12	86	8.5	291	2	US-07-952-277A-7	Sequence 7, Appli

THIS PAGE BLANK (USPTO)


```
Db 1192 AGTGTAGTAAAGAAAGAAATGATAGAGATCGCCAAAGTAGACAATTTATTTGTTTCAAAAT 1251
      ::::: ||| ::::: |||
Qy 176 -----AspLysGluMetThrSerAlaLys 183
      ::::: ||| ::::: |||
Db 1252 GATGGTAATACATTAATACAAATAAAGAGAGATCAAAATCAAAA 1296
      ::::: ||| ::::: |||

RESULT 39
US-09-994-485-7
; Sequence 7, Application US/09994485
; Patent No. US20020142429A1
; GENERAL INFORMATION:
; APPLICANT: Ryazanov, Alexey G.
;            Hait, William N.
;            Pavur, Karen S.
; TITLE OF INVENTION: ELONGATION FACTOR-2 KINASE (EF-2 KINASE)
; AND METHODS OF USE THEREFOR
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
;            Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/994,485
; FILING DATE: 27-NO. US20020142429A1-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/914,999
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 601-1-078
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2237 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEetical: NO
; ORIGINAL SOURCE:
; ORGANISM: Dictyostellium discoideum
; SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-994-485-7

Alignment Scores:
Pred. No.: 0.423 Length: 2237
Score: 87.00 Matches: 45
Percent Similarity: 34.42% Conservative: 29
Best Local Similarity: 20.93% Mismatches: 69
Query Match: 8.62% Indels: 72
DB: 10 Gaps: 11

US-09-868-352-23 (1-192) x US-09-994-485-7 (1-2237)

Qy 12 TyrAsnLysTyrProGly-----ProGln---PheIle 21
      ||||| |||
Db 739 TATAATAATAATAGTGGATTCGTTAGTAATGATGAAGAAATAACACCACAATCATTCCTCT 798
      ||||| |||

Qy 22 HisPheGluAsnIleVallyLysSerAspPheIleGluPheGlnLeuValIleAsnGluLys 41
```

```
Db 799 CATTTTC-----ACATATGAACATTCAAATCATCAATTATTGATT----- 837
      ||||| ::::: ||||| ::::: |||||
Qy 42 SerAlaPheAspValThrValPheGlyGlnArgPheSerGluIleLeuLeu---LysTyr 60
      ||||| ::::: ||| ::::: |||
Db 838 -----ATAGATATTTCACAGGTGTTGGTGATCACTATAGACAGCCCAAAATTCATACCTAT 891
      ||||| ::::: ||| ::::: |||
Qy 61 Asp-----PheIleValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPhe 77
      ||| ::::: ||| ::::: |||
Db 892 GATGCTGTTGGTTTGGTATTGGTAATTGGGTCAA-----AAAGGTTT 936
      ||| ::::: ||| ::::: |||
Qy 78 TyrLysAspAlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIle 97
      ||| ::::: ||| ::::: |||
Db 937 GAAAGAGTTTATAGATACATCAATA----- 960
      ||| ::::: ||| ::::: |||
Qy 98 LysGluTyrCysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArg--- 116
      ||||| ||| ||| |||
Db 961 -----TGTAAATGCAATTTGCCAATATTAAATTTTACAATCAATTAATCCAAATCT 1011
      ||||| ||| ||| |||
Qy 117 -----AspIleLysPheAspGluArgPro 125
      ||||| ||| ||| |||
Db 1012 GAAAAAGTGAATTGTTGTTACTGTACCAAGACCAGATTTAAATTTTCCCTGATACATCTGAA 1071
      ||||| ||| ||| |||
Qy 126 HisLysArgArgLysSerArgSerLysSerGlnSerLysSerGlnThrArgAsnAsn 145
      ::::: ||| ::::: |||
Db 1072 AGAGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 1131
      ::::: ||| ::::: |||
Qy 146 ArgSerGlnSerAsnAlaHisPheThrSerLysLys-----Arg 160
      ||| ::::: ||| ::::: |||
Db 1132 AATAGTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 1191
      ||||| ||| ||| |||
Qy 161 LysAspThrLysArgGlnGluArgHisIleLysGluGluGln----- 175
      ::||| ||| ||| |||
Db 1192 AGTGTAGTAAAGAAAGAAATGATAGATTCGCCAAAGTAGACAATATTATTGTTTCAAAAT 1251
      ::||| ||| ||| |||
Qy 176 -----AspLysGluMetThrSerAlaLys 183
      ::||| ||| ||| |||
Db 1252 GATGGTAATACATTAATAATAAAGAGAGATCAAAATCAAAA 1296
      ::||| ||| ||| |||

RESULT 40
US-09-764-869-1279/c
; Sequence 1279, Application US/09764869
; Patent No. US20020061521A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007
; CURRENT APPLICATION NUMBER: US/09/764,869
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2442
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1279
; LENGTH: 8317
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-869-1279

Alignment Scores:
Pred. No.: 2.73 Length: 8317
Score: 87.00 Matches: 19
Percent Similarity: 50.00% Conservative: 12
Best Local Similarity: 30.65% Mismatches: 29
Query Match: 8.62% Indels: 2
DB: 10 Gaps: 1

US-09-868-352-23 (1-192) x US-09-764-869-1279 (1-8317)

Qy 123 GluArgProHisLysArgArgLysSerArgSerLysSerGlnSerSerLysSerGlnThr 142
      ||||| ::||| ::|||
Db 3152 GAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 3093
      ||||| ::||| ::|||

Qy 143 ArgAsnAsnArgSerGlnSerAsnAlaHisPheThrSerLysLysArgLysAsp 162
```

[illegible]

Ov 161 LvsAspThrLvsArqArqGlnGluArqHisIleLvsGluGln----- 175


```

; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4386
; LENGTH: 609
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-796-692-4386

Alignment Scores:
Pred. No.: 0.0377 Length: 609
Score: 89.00 Matches: 27
Percent Similarity: 48.39% Conservative: 33
Best Local Similarity: 21.77% Mismatches: 50
Query Match: 8.82% Indels: 14
DB: Gaps: 4

US-09-868-352-23 (1-192) x US-09-796-692-4386 (1-609)
Qy 74 LeuArgGlyPheTyrlYsAspAlaSerThrIleArg---LysAsnSerArgIleSerArg 92
|||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 483 CTCAAAGATTTCATGAGCAGCGTGGGGAAGTACCGTTTGGCGATGCACACCGCCTAAA 424

Qy 93 LeuGluAspTyrlleLysGluTyrlCysAsnPheGlyCysAlaTyrlPheValLeuGlu--- 111
|||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 423 TTAATGAAGGGTGGTGGTTGGCTCTTATGGTGACTTAAGAATGCTATTGAAAAA 364

Qy 112 -----AsnProAsnProArgAspIleLys---PheAspAspGluArgProHis 126
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 363 CTTTCTCGAAGGAAATAATGGGAGAAAATAAATTGAAGCGCAGCAAAAGGCAC 304

Qy 127 LysArgArgLysSerArgSerLysSerGlnSer-----SerLys 139
|||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 303 AGTAGGTCAGAGCAGCGTCTCGATCCCGACCAAGAGTTTCCTCTAGGTCTCGTAGCCGA 244

Qy 140 SerGlnThrArgAsnArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysLys 159
|||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 243 TCCCGTTCGGGTAGTCGCAATCTTACAGCGGTCAAGAGCAGGAGCGGAGC 184

Qy 160 ArgLysAspThrLysArgArgGlnGluArgHisIleLysGluGlnAspLysGluMet 179
|||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 183 CGGAGCAAGTCCCGTCTGTAGTAGGTCTCCCGTCCCTGAGAAGAGCCAGAACGTGGT 124

Qy 180 ThrSerAlaLys 183
:::||||:::
Db 123 TCTTCAAGTAGA 112

RESULT 34
US-09-835-232-7/c
; Sequence 7, Application US/09835232
; Patent No. US20020098489A1
; GENERAL INFORMATION:
; APPLICANT: Leder, Philip
; APPLICANT: Leader, Benjamin
; TITLE OF INVENTION: FORMIN-2 NUCLEIC ACIDS AND POLYPEPTIDES
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 00383/052002
; CURRENT APPLICATION NUMBER: US/09/835,232
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: US 60/196,811
; PRIOR FILING DATE: 2000-04-13
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 170834
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(170834)
; OTHER INFORMATION: n= A,T,C, or G
US-09-835-232-7

Alignment Scores:
Pred. No.: 151 Length: 170834

```



```
; LENGTH: 7521
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(7521)
US-09-819-104A-3

Alignment Scores:
Pred. No.: 1,16 Length: 7521
Score: 89.50 Matches: 29
Percent Similarity: 40.85% Conservative: 29
Best Local Similarity: 20.42% Mismatches: 53
Query Match: 8.87% Indels: 31
DB: 9 Gaps: 5

US-09-868-352-23 (1-192) x US-09-819-104A-3 (1-7521)
QY 72 LeuArgLeuArgGlyPheTyrLysAspAlaSerThrIleArgLysAsnSerArgIleSer 91
Db 1225 ATCAACATGAACGGCTTATGGCCGACCCCATGAAGGTGTACAAAGACCCCGAGTCATG 1284
QY 92 ArgLeu-----GluAspTyrIleLysGluTyrCys 101
Db 1285 AACATGTGGAGTCAGCAGGAGAGACCTTCCGGGAGAGTTTCATGCAGCATCCCAAG 1344
QY 102 AsnPheGly-----CysAla----- 106
Db 1345 AACTTTGGCCTGATCGCATCTCTGGAGAGGACAGTGGCTGAGTGGCTCTCTAT 1404
QY 107 ---TyrPheValLeuGluAsnProAsnProArgAspIleLysPheAspGluArgPro 125
Db 1405 TACTACTGCTACTAAGAAGATGAGAACTATAGAGCCTG-----GTGAGACGG 1452
QY 126 HisLysArgArgLysSerArgSerLysSerGlnSerSerLysSerGlnThrArgAsnAsn 145
Db 1453 AGCTATCGGCGCGCGAGAGAGCCAGCACACAGCAGCAGCAGCAGCAGCAGCAG 1512
QY 146 ArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysLysArgLysAspThrLysArg 165
Db 1513 CAGCAGCAGCAGCAGCCATGCCCGC---AGCAGCAGGAGGAGAAAGATGAGAAGGAG 1569
QY 166 ArgGlnGluArgHisIleLysGluGluGlnAspLysGluMetThrSerAlaLysGlnHis 185
Db 1570 AAGAAAGAGCGGAGAGAGGAGGAGAGAGCCGGAGTGGAGACGACGACGAGGAGAC 1629
QY 186 LeuLeu 187
Db 1630 CTCCTC 1635

RESULT 27
US-09-819-104A-1
; Sequence 1, Application US/09819104A
; Publication No. US20030027137A1
; GENERAL INFORMATION:
; APPLICANT: Chen, J. Don
; TITLE OF INVENTION: NOVEL NUCLEAR RECEPTOR COREPRESSOR MOLECULES
; FILE REFERENCE: UM-030
; CURRENT APPLICATION NUMBER: US/09/819,104A
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: 60/193,138
; PRIOR FILING DATE: 2000-03-29
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 8686
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (157)..(7677)
US-09-819-104A-1
```

```
Alignment Scores:
Pred. No.: 1,43 Length: 8686
Score: 89.50 Matches: 29
Percent Similarity: 40.85% Conservative: 29
Best Local Similarity: 20.42% Mismatches: 53
Query Match: 8.87% Indels: 31
DB: 9 Gaps: 5

US-09-868-352-23 (1-192) x US-09-819-104A-1 (1-8686)
QY 72 LeuArgLeuArgGlyPheTyrLysAspAlaSerThrIleArgLysAsnSerArgIleSer 91
Db 1381 ATCAACATGAACGGCTTATGGCCGACCCCATGAAGGTGTACAAAGACCCCGAGTCATG 1440
QY 92 ArgLeu-----GluAspTyrIleLysGluTyrCys 101
Db 1441 AACATGTGGAGTCAGCAGGAGAGACCTTCCGGGAGAGTTTCATGCAGCATCCCAAG 1500
QY 102 AsnPheGly-----CysAla----- 106
Db 1501 AACTTTGGCCTGATCGCATCTCTGGAGAGGACAGTGGCTGAGTGGCTCTCTAT 1560
QY 107 ---TyrPheValLeuGluAsnProAsnProArgAspIleLysPheAspGluArgPro 125
Db 1561 TACTACTGCTACTAAGAAGATGAGAACTATAGAGCCTG-----GTGAGACGG 1608
QY 126 HisLysArgArgLysSerArgSerLysSerGlnSerSerLysSerGlnThrArgAsnAsn 145
Db 1609 AGCTATCGGCGCGCGAGAGAGCCAGCACACAGCAGCAGCAGCAGCAGCAGCAG 1668
QY 146 ArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysLysArgLysAspThrLysArg 165
Db 1669 CAGCAGCAGCAGCAGCCATGCCCGC---AGCAGCAGGAGGAGAAAGATGAGAAGGAG 1725
QY 166 ArgGlnGluArgHisIleLysGluGluGlnAspLysGluMetThrSerAlaLysGlnHis 185
Db 1726 AAGAAAGAGCGGAGAGAGGAGGAGAGCCGGAGTGGAGACGACGACGAGGAGAC 1785
QY 186 LeuLeu 187
Db 1786 CTCCTC 1791

RESULT 28
US-09-960-352-5438/c
; Sequence 5438, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 5438
; LENGTH: 203
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 23-LIB34-075-Q1-E1-F11
US-09-960-352-5438

Alignment Scores:
Pred. No.: 0,00791 Length: 203
Score: 89.00 Matches: 13
Percent Similarity: 62.96% Conservative: 21
Best Local Similarity: 24.07% Mismatches: 20
Query Match: 8.82% Indels: 0
DB: 10 Gaps: 0
```



```
Db 654 ATTAGGCTTATTGAAGATAAGCCACGACAGCATAGGCGATCTTACTCTGGAAGCAGA 713
QY 126 -----HisLysArgArgLysSerArgSerLysSer---GlnSerSerLysSerGln 141
Db 714 TCCAGGTCTCGATCTAGAACACGGTCACGAAGTAGGAGTCGACGAGCAGCCGCGAGTAGA 773
QY 142 ThrArgAsn---AsnArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysLysArg 160
Db 774 TCTCGAAGTATCTCAAAAGTCGCTCCGTTCCAGGTCGCGGAGCAAGGTCGATCAGCT 833
QY 161 LysAspThrLysArgArgGlnGluArgHisIleLysGluGluGlnAspLysGluMetThr 180
Db 834 TCTCGATCAAAAGCAGCAAAATCTAGATCAAAAGCAAAATCTAAGCCCAAGTCTGATCGG 893
QY 181 SerAlaLysGlnHis 185
Db 894 GGTCTCCATTCACAT 908

RESULT 21
US-09-728-952-78
; Sequence 78, Application US/09728952
; Patent No. US20020111302A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Goodrich, Ryle
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Wang, Dunrui
; APPLICANT: Yamazaki, Vicki
; APPLICANT: Ujwal, Manusha L.
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20020111302A1el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 799
; CURRENT APPLICATION NUMBER: US/09/728,952
; CURRENT FILING DATE: 2000-11-30
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: pt_FL-genes Version 2.0
; SEQ ID NO 78
; LENGTH: 5273
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (93)..(4832)
US-09-728-952-78

Alignment Scores:
Pred. No.: 0-527 Length: 5273
Score: 90.50 Matches: 46
Percent Similarity: 34.90% Conservative: 21
Best Local Similarity: 23.96% Mismatches: 62
Query Match: 8.97% Indels: 63
DB: 10 Gaps: 7

US-09-868-352-23 (1-192) x US-09-728-952-78 (1-5273)
QY 24 GluAsnIleValLysSerAspAlleGluPheGlnLeuValIleAsnGluLysSerAla 43
Db 4278 AAAAATATTGTTCAA---GATGACAAAGAGGTGATCTTGAGCTCAGAGGAGGAGTTTC 4334
QY 44 PheAspValThrValPheGlyGlnArgPheSerGluIleLeuLysTyrAspPheIle 63
Db 4335 TTT-----GTCCAA 4343
QY 64 ValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAspAlaSerThr 83
Db 4344 GTCCATGATGTTTCTCCAGAGCAACCTCGA-----ACA 4376
QY 84 IleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGlu---TyrCysAsn 102
Db 4375 GTCATCAAAAGCAGCCCGCGTCCAGCACTGCAGAGGATGTCTATTCCAGCAGACCTTTATGCAA 5034
```

```
Db 4377 GTCATCAAAAGCAGCCCGCGTCCAGCACTGCACAGGATGTCATTTCAGCAGACCTTTATGCAA 4436
QY 103 PheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspIleLysPheAspAsp 122
Db 4437 GCCAAATATTCCTACAGCATCTCTGAGCAGCAGCCCAATCCCAAGCGACTATGCTTTTGGAA 4496
QY 123 Glu-----ArgProHisLysArgArgLysSerArgSerLysSerGln----- 136
Db 4497 GAGGTGGTGAAGACACTACCAACAAGAGACTACCACACCAAGTCTCTCAGCGGGTC 4556
QY 136 ----- 136
Db 4557 CTTCTGCAGCAGGTGTGTGTTTCAAGCCCAAGGGAAGGTGGAGAAAGTGCAGGAAATTC 4616
QY 137 -----SerSerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHis 154
Db 4617 ATCCTTAAGCTAAAGGAGCAGGTGCGAGGCATCTCGAGAGATATAAAAGAAAGCATTTCT 4676
QY 155 PheThrSerLysLysArgLysAspThrLysArgArgGlnGluArgHisIleLysGluGlu 174
Db 4677 TTGCAAGTGAAGTCAAGAGCTCACCAG-----TCAACTAAA 4715
QY 175 GlnAspLysGluMetThrSerAlaLysGlnHisLeu 186
Db 4716 CAGCCCGAGGAGCTTACATCACCTTCTCAGCTCTTG 4751

RESULT 22
US-09-822-635-3
; Sequence 3, Application US/09822635
; Patent No. US20010039331A1
; GENERAL INFORMATION:
; APPLICANT: Hunter, John J.
; APPLICANT: Meyers, Rachel A.
; TITLE OF INVENTION: 16836, A NOVEL HUMAN PHOSPHOLIPASE C AND
; FILE REFERENCE: 10448-035001
; CURRENT APPLICATION NUMBER: US/09/822,635
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: US 60/193,921
; PRIOR FILING DATE: 2000-03-31
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 5430
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-635-3

Alignment Scores:
Pred. No.: 0-55 Length: 5430
Score: 90.50 Matches: 46
Percent Similarity: 34.90% Conservative: 21
Best Local Similarity: 23.96% Mismatches: 62
Query Match: 8.97% Indels: 63
DB: 10 Gaps: 7

US-09-868-352-23 (1-192) x US-09-822-635-3 (1-5430)
QY 24 GluAsnIleValLysSerAspAlleGluPheGlnLeuValIleAsnGluLysSerAla 43
Db 4876 AAAACATTTGTTCAA---GATGACAAAGAGGTGATCTTGAGCTCAGAGGAGGAGTTTC 4932
QY 44 PheAspValThrValPheGlyGlnArgPheSerGluIleLeuLysTyrAspPheIle 63
Db 4933 TTT-----GTCCAA 4941
QY 64 ValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAspAlaSerThr 83
Db 4942 GTCCATGATGTTTCTCCAGAGCAACCTCGA-----ACA 4974
QY 84 IleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGlu---TyrCysAsn 102
Db 4975 GTCATCAAAAGCAGCCCGCGTCCAGCACTGCAGAGGATGTCTATTCCAGCAGACCTTTATGCAA 5034
```

```

Qy   174 GluClnAspLysGluMetThrSerAlaLys 183
      :::                ::::::::::|||
Db    212 AGGGTGAAGAAGGTCAAGGACGCGTCAAAA 241

RESULT 20
US-10-002-600-58
; Sequence 58, Application US/10002600
; Patent No. US20020137077A1
; GENERAL INFORMATION:
; APPLICANT: Hopkins, Christopher M.
; APPLICANT: Peterson, David P.
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Hawkins, Phillip R.
; TITLE OF INVENTION: GENES REGULATED IN ACTIVATED T CELLS
; FILE REFERENCE: PA-0042 US
; CURRENT APPLICATION NUMBER: US/10/002,600
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: 60/243,521
; PRIOR FILING DATE: 2000-10-25
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: PERL Program
; SEQ ID NO 58
; LENGTH: 3697
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Template ID: 039976.16
; NAME/KEY: unsure
; LOCATION: 1444, 1446, 1448-1449, 1454, 1456
; OTHER INFORMATION: a, t, c, g, or other
US-10-002-600-58

Alignment Scores:
Pred. No.:          0.276           Length:       3697
Score:             91.00           Matches:      52
Percent Similarity: 38.22%         Conservative: 34
Best Local Similarity: 23.11%      Mismatches:   83
Query Match:        9.02%          Indels:       56
DB:                 12            Gaps:         9

US-09-868-352-23 (1-192) x US-10-002-600-58 (1-3697)

Qy   17 GlyProGlnPheIleHisPheGluAsnIleValLysSerAspSpilleGluPheGlnLeu 36
      |||               ||||||:::  ::|||::|  :|||::|  |::|::|
Db    234 GGGTACGCCTTCGTGGAGTTCCGAGGACTCCCGGACGCCGACGACGCCGTTTACGAGCTG 293
      |||               |||         |||         -----PheGlyGlnArgPhe 53
Qy   37 ValIleAsnGluLysSerAlaPheAspValThrVal-----PheGlyGlnArgPhe 53
      |||               |||         |||         |||         |||
Db    294 AACGGCAAGGAGCTCTCGCGCAGCGCGTGATCGTAGAGCACGCCCGGCGCGCGCTCGC 353
      |||               |||         |||         |||         |||
Qy   54 SerGluIleLeuLeuLysTyrAspPheIleValGlyAspTrpGlyAsnGluGlnLeuArg 73
      |||               |||         |||         |||         |||
Db    354 GATCGCAGCGCTACAGCTACGGAGCGCGAGTGGTGGAGTGGATACAGCAGCTCGGAGA 473
      |||               |||         |||         |||         |||
Qy   74 LeuArgGlyPheTyrlLysAspAlaSerThriLeArgLysAsnSerArgIle----- 90
      |||               |||         |||         |||         |||
Db    414 ACATCTGGCAGACAATAACGGACCACCTGTTTCGTACAGAATACAGGCTTATTGTAGAA 473
      |||               |||         |||         |||         |||
Qy   91 -----SerArg-----LeuGluAspTyrlle----- 97
      |||               |||         |||         |||         |||
Db    474 AATCTTCTAGTCGTCAGTTGGCAAGATTAAAGGATTTTATGGCAACAGCAGGTGAA 533
      |||               |||         |||         |||         |||
Qy   98 -----LysGluTyrcysAsnPheGlycysAlaTyrlphe----- 108
      |||               |||         |||         |||         |||
Db    534 GTAACCTATCGGATGCCACAGGAAACGAACAAATAGGGTGTAATGTAGTTTCGCTCC 593
      |||               |||         |||         |||         |||
Qy   109 -----ValLeuGlu-----AsnProAsnProArgasp 117
      |||               |||         |||         |||         |||
Db    594 TACTCTCATGACGCTGCTTTGGCAAACTGGATGGCACAAGAAATAAATGCGCAAAAT 653
      |||               |||         |||         |||         |||
Qy   118 IleLysPheAspGspGluArgPro----- 125
      |||               |||         |||         |||         |||

```

```

: PRIOR APPLICATION NUMBER: PCT/US01/005653
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/005662
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/005661
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/005670
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: US 60/234,687
: PRIOR FILING DATE: 2000-09-21
: PRIOR APPLICATION NUMBER: US 09/608,408
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: US 09/774,203
: PRIOR FILING DATE: 2001-01-29
: NUMBER OF SEQ IDS: 49117
: SOFTWARE: Anomax Sequence Listing Engine
: SEQ ID NO 21362

```

Alignment Scores:		
Pred. No.:	0.0447	Length:
Score:	91.50	Matches:
Percent Similarity:	40.35%	Conservative:
Best Local Similarity:	28.07%	Mismatches:
Query Match:	9.07%	Indels:
DB:	10	Gaps:
		4
US-09-868-352-23 (1-192)	x US-09-864-761-21362 (1-1134)	

[illegible]

Db 221 AATGCCAAAGACCTTTGTGTGAGCGAGTAATTGTTGACATGCCCGCGGCCACGGGCA 280
Qy 37 -----ValIleAsnGluLysSerAlaPheAsp 45
Db 281 GATGCCAGTTACGGTTCTGGACGCATGATATGGTTATAGAAAGATGCCGAGATAAAT 340
Qy 46 ValThrValPheGlyGlnArgPheSerGluIleLeuLeuLysTyrAspPheIleValGly 65
Db 341 ATGGCCCTCTACTCGCACAGAGTACAGACTTATTGTGGAGAATTTGTCAAGTCGG-TGC 399
Qy 66 AspTrpGlyAsn-----GluGlnLeuArgLeuArgGly-----PheTyrLysAspAla 81
Db 400 AGCTGGCAAGACCTTAAGGATTATATGGCTCAGGCAGGAGAGTGACTTATGCAGATGCT 459
Qy 82 SerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyrCys 101
Db 460 CACAAGGGACGCGCAAAAT-----GAAGGGGTGATTGAATTTGTA 498
Qy 102 AsnPheGlyCysAlaTyrPheValLeuGlu-----AsnProAsnProArg 116
Db 499 TCTTATTCTGATATAAAGAGCTTTGGAAAGTTGGATGGAACCTGAAGTCATCGGAGA 558
Qy 117 AspileLysPheAspAspGluArgPro-----ArgLysSerArgSerLysSer 125
Db 559 AAAATCAGATTAGTTGAAGACAGCCAGGTTCCAGACGACGCCGCTCCTACTCCAGAAGC 618
Qy 126 -----HisLysArg-----ArgLysSerArgSerLysSer 135
Db 619 CGGAGTCATTCAGGTCGCTCTCGAGCAGACATTCCTGTAAGAGCAGGAAGCCGAAGT 678
Qy 136 GlnSerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPhe 155
Db 679 GGCACACCAAGACAGTCATCTAAGAGTAGATCTCGGTCCAGGTCGGGCTCCCGCTCC 738
Qy 156 ThrSerLys-----LysArgLysAspThrLysArgArgGlnGluArgHisIleLysGlu 173
Db 739 CGGACCAAGCGCGGAGCGGAGCCAGAGTCGGAGCGGAGCAAGAAAGAGAAAAGCAGG 798
Qy 174 GluGlnAspLysGluMetThrSerAlaLysGlnHis 185
Db 799 ACCCCAGCAGGACAGAGCGGACCGCGAGCCGACCAT 834
RESULT 13
; Sequence 71, Application US/10001843
; Patent No. US20020132255A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Cafferkey, Robert
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; APPLICANT: Turner, Leah
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and Pr
; FILE REFERENCE: DEX-0267
; CURRENT APPLICATION NUMBER: US/10/001,843
; PRIOR FILING DATE: 2001-11-20
; PRIOR FILING DATE: 2000-11-20
; NUMBER OF SEQ ID NOS: 218
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 71
; LENGTH: 3259
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-001-843-71
Alignment Scores:
Pred. No.: 0.0639 Length: 3259
Score: 95.50 Matches: 19
Percent Similarity: 64.91% Conservative: 18
Best Local Similarity: 33.33% Mismatches: 15

Query Match: 9.46% Indels: 5
Db: 12 Gaps: 1
US-09-868-352-23 (1-192) x US-10-001-843-71 (1-3259)
Qy 121 AspAspGluArgProHisLysArgArgLysSerArgSerLysSerGlnSerSerLysSer 140
Db 3051 AACGAAGAGAGAACGACGAAGAAAGAAAGACAGCAGCAAGAAAGAGAGAAAGAGA 3110
Qy 141 GlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysLysArg 160
Db 3111 AGAAGCAAGAAAGACAGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 3155
Qy 161 LysAspThrLysArgArgGlnGluArgHisIleLysGluGlnAspLys 177
Db 3156 CGGAAACACGACGAAGAGAAAGAAAGAAAGAGAGAGAGAGAGAGAGAGAG 3206
RESULT 14
US-09-989-920-113
; Sequence 113, Application US/09989920
; Patent No. US20020172957A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Chen, Sei-Yu
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and P
; FILE REFERENCE: DEX-0291
; CURRENT APPLICATION NUMBER: US/09/989,920
; CURRENT FILING DATE: 2001-11-21
; PRIOR FILING DATE: 60/252,500
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 284
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 113
; LENGTH: 2269
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2266)..(2266)
; OTHER INFORMATION: a, c, g or t
US-09-989-920-113
Alignment Scores:
Pred. No.: 0.0676 Length: 2269
Score: 93.50 Matches: 19
Percent Similarity: 55.56% Conservative: 16
Best Local Similarity: 30.16% Mismatches: 27
Query Match: 9.27% Indels: 1
Db: 9 Gaps: 1
US-09-868-352-23 (1-192) x US-09-989-920-113 (1-2269)
Qy 123 GluArgProHisLysArgArgLysSerArgSerLysSerGlnSerSerLysSerGlnThr 142
Db 1877 AACACGA 1936
Qy 143 ArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLys---LysArgLys 161
Db 1937 AAAAGACAAAGAGAAAGAAACACAGAAAGAAACCAACAAACAAAGAGAGAGAGAG 1996
Qy 162 AspThrLysArgArgGlnGluArgHisIleLysGluGlnAspLysGluMetThrSer 181
Db 1997 GAAAGAGAGAGACGCAAGAGAGAGAAACAAAGAAAGAGAGAGAGAGAGAGAGAG 2056
Qy 182 AlaLysGln 184
Db 2057 GAGAGAAGA 2065
RESULT 15
US-09-878-574-4310/c


```
; Sequence 1933, Application US/09867550
; Patent No. US20020082206A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Mehraban, Fuad,
; APPLICANT: Conley, Pamela
; APPLICANT: Law, Debbie
; APPLICANT: Topper, James
; TITLE OF INVENTION: NO. US20020082206A1el Polynucleotides from Atherogenic Cells and
; FILE REFERENCE: 21402-013 (Cura-313)
; CURRENT APPLICATION NUMBER: US/09/867,550
; CURRENT FILING DATE: 2001-09-20
; PRIOR APPLICATION NUMBER: USSN 60/208,427
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 2125
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1933
; LENGTH: 712
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-867-550-1933

Alignment Scores:
Pred. No.: 0.00271 Length: 712
Score: 99.00 Matches: 29
Percent Similarity: 46.00% Conservative: 17
Best Local Similarity: 29.00% Mismatches: 22
Query Match: 9.81% Indels: 32
DB: 10 Gaps: 4

US-09-868-352-23 (1-192) x US-09-867-550-1933 (1-712)

Qy 115 ProArgAspIleLysPheAspAspGluArgProHis----- 126
Db 271 CCTCGTTCACATCTTATGATAGAGAGCGAGCATCGATCAAGCAGTAGCTCTTCTTAT 330
Qy 127 -----LysArgArgLysSerArgSerLysSerLysSer----- 137
Db 331 GCCTCCAGAGAAACGAAGTCGAAGTCGTTCAAGGGGTGAGGAAATCCTATAGAGTT 390
Qy 138 -----SerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAla 153
Db 391 CAGAGGTCCTAGGTCAAAAAGCAGCAAGAGGTCCAGGTCAAGACCTCGTCTCGTCTCT 450
Qy 154 HisPheThrSerLysLysArg-----LysAspThr 163
Db 451 CATAGCTGATGAGTGAAGGTCCAGTCACAGAAAGCGGTAGTCGTCTCGGATAGA 510
Qy 164 LysArgArgGlnGluArgHisIleLysGluGlnAspLysGluMetThrSerAlaLys 183
Db 511 GAACGACGTAAAGGCAGAGAT---AAAGAGAAAGAGAAAGGAGAGATAAAGGGAG 567

RESULT 8
US-09-925-301-557
; Sequence 557, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 557
; LENGTH: 3484
; TYPE: DNA
; ORGANISM: Homo sapiens
```

```
US-09-925-301-557

Alignment Scores:
Pred. No.: 0.0259 Length: 3484
Score: 99.00 Matches: 24
Percent Similarity: 58.57% Conservative: 17
Best Local Similarity: 34.29% Mismatches: 21
Query Match: 9.81% Indels: 8
DB: 10 Gaps: 2

US-09-868-352-23 (1-192) x US-09-925-301-557 (1-3484)

Qy 122 AspGluArgProHisLysArgLysSerArgSerLysSerGlnSer----- 137
Db 1101 GATAGAAAGAGAAAAAGAGGCATTCAAGATCAAGTACACGTCTTAGGAGGAGGAGG 1160
Qy 138 -----SerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPhe 155
Db 1161 ACTCCCTCATCTTCTAGACACAGCGGCTCAAGAAGCAGATCGAGACGGCGGTCACATTTCT 1220
Qy 156 ThrSerLysLysArgLysAspThrLys-----ArgArgGlnGluArgHisIleLysGlu 173
Db 1221 AAGTCTAGGATCGGCGCAGCATCCAAAAGCCCAAGCGGAGAAATCTCATTTCCAGAGAA 1280
Qy 174 GluGlnAspLysGluMetThrSerAlaLys 183
Db 1281 AGAGGTAGAGGTCAAGGAGGCACATCAAAA 1310

RESULT 9
US-09-834-975-822
; Sequence 822, Application US/09834975
; Patent No. US20020110815A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Brown, Jeffrey
; APPLICANT: Bolt, Andrew
; APPLICANT: Van Huffel, Christophe
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS AND METHODS
; TITLE OF INVENTION: FOR THE IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; FILE REFERENCE: MRI-016B
; CURRENT APPLICATION NUMBER: US/09/834,975
; CURRENT FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/197,538
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 1046
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 822
; LENGTH: 2025
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(2025)
; OTHER INFORMATION: n = A,T,C or G

US-09-834-975-822

Alignment Scores:
Pred. No.: 0.0184 Length: 2025
Score: 97.50 Matches: 19
Percent Similarity: 52.86% Conservative: 18
Best Local Similarity: 27.14% Mismatches: 32
Query Match: 9.66% Indels: 1
DB: 10 Gaps: 1

US-09-868-352-23 (1-192) x US-09-834-975-822 (1-2025)

Qy 116 ArgAspIleLysPheAspAspGluArgProHisLysArgArgLysSerArgSerLysSer 135
Db 1587 CAGAACGCCCAACGACGACGAGCAAAAAGCCGAGAGAAAGAGAGAGAGAGAGAGAGAG 1646
Qy 136 GlnSerSerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPhe 155
Db 136 GlnSerSerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPhe 155
```

```

Db 1 ATGGTGACGATTTTCAGCATGTAATGAAGTCTCGAAACAGCGCANTGTTGGAAT 60
Qy 46 ValThrValPheGlyGlnArgPheSerGluIleLeuLeuLysTyrAspPheIleValGly 65
Db 61 GAGGAGCATTCAAAGCGGTACAGCGTACTGAACAAATATGATACATCGTCGGC 120
Qy 66 AspTrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAsp-----AlaSer 82
Db 121 GATTTGGGCTACAAACCACTGCGGTTGCGCGGCTTTTTCATGACACAAACAAAAATCA 180
Qy 83 ThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyrCysAsn 102
Db 181 ACATTGCGATAAA-----AAATAGACGCTTACCCGACTATCTTTACGAATACTGCAAT 234
Qy 103 PheGlyCysAlaTyrPheValLeu 110
Db 235 TTTGGTTGCGCTTATTTGTCITG 258

RESULT 5
US-09-974-300-3145
; Sequence 3145, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3145
; LENGTH: 273
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-3145

Alignment Scores:
Pred. No.: 3.48e-19 Length: 273
Score: 222.50 Matches: 42
Percent Similarity: 74.68% Conservative: 17
Best Local Similarity: 53.16% Mismatches: 19
Query Match: 22.05% Indels: 1
DB: 10 Gaps: 1

US-09-868-352-23 (1-192) x US-09-974-300-3145 (1-273)
Qy 34 PheGlnLeuValIleAsnGluLysSerAlaPheAspValThrValPheGlyGlnArgPhe 53
Db 26 TTTGAAGTCGTCAAGATGTGAAGAAAGCGTTTAAACGAAGAACGCTTTAAAGCCAGATAC 85
Qy 54 SerGluIleLeuLeuLysTyrAspPheIleValIclyAspTrpGlyAsnGluGlnLeuArg 73
Db 86 TCCGACATTTTAAATAATACGACTACATCGTCGGGATTGGGGATACACCCAGCTGAGS 145
Qy 74 LeuArgGlyPheTyrLysAspAlaSer---ThrIleArgLysAsnSerArgIleSerArg 92
Db 146 CTGAAGCGCTTTTTCAGCATCAAAATCAAAAGGCCACATTTGTATACGAAGATCAGCACG 205
Qy 93 LeuGluAspTyrIleLysGluTyrCysAsnPheGlyCysAlaTyrPheValLeuGlu 111
Db 206 CTTGATGAATATATTACGAGTACTGCAATTCGGATGTGCCACTTTGTTTAAAAA 262

RESULT 6
US-10-044-090-268
; Sequence 268, Application US/10044090
; Patent No. US20020137081A1
```

```

; GENERAL INFORMATION:
; APPLICANT: Olga Bandman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044,090
; CURRENT FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 268
; LENGTH: 1534
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020137081A1 988704.26
US-10-044-090-268

Alignment Scores:
Pred. No.: 0.007 Length: 1534
Score: 99.50 Matches: 43
Percent Similarity: 41.23% Conservative: 44
Best Local Similarity: 20.38% Mismatches: 73
Query Match: 9.86% Indels: 51
DB: 12 Gaps: 8

US-09-868-352-23 (1-192) x US-10-044-090-268 (1-1534)
Qy 20 PheIleHisPheGluAsnIleValLysSerAspIleGluPheGlnLeuValIleAsn 39
Db 293 TTTGTGGAATTTGAGGATCCAGGGATGCAGATGATGCTGTGTATGAGCTTGAGAAAA 352
Qy 40 GluLysSerAlaPheAspValThrVal----- 48
Db 353 GAACCTGTGTAGTCAAAAGGGTTACTATTGAAACATGCTAGGCTCGGTACAGAGGTGGAAGA 412
Qy 49 -----PheGlyGlnArgPheSerGluIleLeuLysTyrAsp----- 61
Db 413 GGTAGAGGAGGATACTCTGACCGTTTTAGTAGTCGCGAGACCTCGAAATGATAGACGAAAT 472
Qy 62 -----PheIleValGlyAsp----- 66
Db 473 GCTCCACCTCTAAGAACACAGAAATCGTCTTATAGTTGAGAAATTTATCTCAAGAGTCAGC 532
Qy 67 TrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAspAlaSerThrIleArg--- 85
Db 533 TGGCAGGAT-----CTCAAAGATTTCATGAGACAAAGCTGGGGAAGTAACGTTT 580
Qy 86 LysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyrCysAsnPheGlyCys 105
Db 581 GCGGATGCACACGACCTAAATTTAAATGAAGGGTGTGAGTTGCGCTCTTATGTTGAC 640
Qy 106 AlaTyrPheValLeuGlu-----AsnProAsnProArgAspIleLys--- 119
Db 641 TTTAAAGATGCTATTGAAACATCTTCTGGAAAGCAATAAATCGGAGAAAAATAAATA 700
Qy 120 PheAspAspGluArgProHisLysArgArgLysSerArgSerLysSerGlnSer----- 137
Db 701 ATTGAAGGACGACAAAGGACACAGTAGGTCAAGAAAGCAGGTCTCGATCCCGACGAGAGT 760
Qy 138 -----SerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAlaAsn 152
Db 761 TCCTCTAGTGTCTCGTAGCCGATCCCGTTCCCGTAGTCGCAAAATCTTACAGCCCGTCAAGA 820
Qy 153 AlaHisPheThrSerLysLysArgLysAspThrLysArgArgGlnGluArgHisIleLys 172
Db 821 ACAGAGGACGAGGACCGGACCGACCAAGTCCCGTCTCTTAGTAGTCTCCCGTGCCT 880
Qy 173 GluGluGlnAspLysGluMetThrSerAlaLys 183
Db 881 GAGAAGACCCAGAAACCGTGGTTCTTCAAGTAGA 913

RESULT 7
US-09-867-550-1933
```

```
Qy 141 GlnThrArgAsnArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysLysArg 160
    |||||
Db 415 CCGACTCAAACTAAACGA--AAGTCGATGCAGAT-----447
Qy 161 LysAspThrLysArgGlnGlnArgHisLysLysGluGlnAspLysGluMetThr 180
    |||
Db 448 AAGAAATAGACGTCGTCAGAAAGACACGACGATTCACAGAAAGAGACAGGAA-----501
Qy 181 SerAlaLysGlnHis-LeuLeuPheValArg 190
    |||||
Db 502 -----CAACGTCATTTGTTCATTCGTCAGCA 526

RESULT 3
US-09-070-927A-502
; Sequence 502, Application US/09070927A
; Patent No. US20020120116A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Kunsch
;           Patrick J. Dillon
;           Steven Barash
; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 982
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070, 927A
; FILING DATE: 04-May-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/046,655
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: 60/044,031
; FILING DATE: 1997-05-06
; APPLICATION NUMBER: 60/066,009
; FILING DATE: 1997-11-14
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB369
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 502:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1495 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 502:
US-09-070-927A-502

Alignment Scores:
Pred. No.: 4,77e-28 Length: 1495
Score: 302.50 Matches: 69
Percent Similarity: 57.74% Conservative: 28
Best Local Similarity: 41.07% Mismatches: 42
Query Match: 29.98% Indels: 29
DB: 10 Gaps: 7

US-09-868-352-23 (1-192) x US-09-070-927A-502 (1-1495)
Qy 26 IleValLysSerAspAspIle-----GluPheGlnLeuValIleAsnGlu 40
    |||||
    |||||
```

```
Db 153 ATTGTGAATGAACAGATATTCAAATTTGCTGAAGAGAGTAGTACCGAATTTGTTAAAAATCAT 212
Qy 41 LysSerAlaPheAspValThrValPheGlyGlnArgPheSerGluIleLeuLeuLysTyr 60
    :::: |||||
Db 213 CGTGATCGCTTTGATCTGAACGCTTAGGTGAACGGTTTACTGAAGTATTAGCTCGTTAT 272
    :::: |||||
Qy 61 AspPheIleValGlyAspTrpGlyAsnGlnLeuArgLeuArgGlyGlyPheThrLysAsp 80
    :::: |||||
Db 273 GATTATATCGTAGGAGACTGGGGCTATGAACAGTTACGCTTGGCGGCTTTT---GAT 329
    :::: |||||
Qy 81 AlaSerThrIleArg-----LysAsnSerArgIleSerArgLeuGluAspTyrIleLys 98
    :::: |||
Db 330 ACTTCTAATCGTCGGCAGCGCGCATCAACGCTTGTATACACTAGAGAGACTATTGTAT 389
    :::: |||
Qy 99 GluTyrCysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspIle 118
    :::: |||||
Db 390 GAATATTGTAATTTGGCTGTCCCTATTTTGTAATTGAACGACGAGGGAACGT-----443
    :::: |||||
Qy 119 LysPheAspGluArgProHisLysArgArgLysSerArgSerLysSerGlnSerSer 138
    :::: |||||
Db 444 -----AAGAAAAAACAACCGCATCGTCGTCGCCATAAA-----479
    :::: |||||
Qy 139 LysSerGlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLys 158
    :::: |||||
Db 480 -----AAAAATAATCGTTCTAACGCG-----TTTGTTCAGGAA 512
    :::: |||||
Qy 159 LysArgLysAspThrLysArgArgGlnGlnArgHisIle-----LysGluGlnAsp 176
    :::: |||
Db 513 AAGAGTCGCCCGCTAATAAATAATAGCAAGCCAGCTGATTCGTACGAGAAAAAGAAAT 572
    :::: |||||
Qy 177 LysGluMetThrSerAlaLysGln 184
    :::: |||||
Db 573 AAAAAACGACTACACCCAAACAG 596

RESULT 4
US-09-974-300-7936
; Sequence 7936, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berk, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7936
; LENGTH: 285
; TYPE: DNA
; ORGANISM: Bacillus clausii
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(285)
; OTHER INFORMATION: n = A,T,C or G
US-09-974-300-7936

Alignment Scores:
Pred. No.: 2.13e-20 Length: 285
Score: 232.50 Matches: 46
Percent Similarity: 69.32% Conservative: 15
Best Local Similarity: 52.27% Mismatches: 22
Query Match: 23.04% Indels: 5
DB: 10 Gaps: 2

US-09-868-352-23 (1-192) x US-09-974-300-7936 (1-285)
Qy 26 IleValLysSerAspAspIleGluPheGlnLeuValIleAsnGluLysSerAlaPheAsp 45
    |||||
```

;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; NUMBER OF SEQ ID NOS: 14110
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 9316
;; LENGTH: 531
;; TYPE: DNA
;; ORGANISM: Streptococcus pneumoniae
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1)...(531)
US-09-815-242-9316

Alignment Scores:
Pred. No.: 2,86e-47 Length: 531
Score: 452.50 Matches: 94
Percent Similarity: 66.49% Conservatives: 33
Best Local Similarity: 49.21% Mismatches: 48
Query Match: 44.85% Indels: 16
DB: 10 Gaps: 4

US-09-868-352-23 (1-192) x US-09-815-242-9316 (1-531)

QY 1 MetArgLysGluValThrProGluMetLeuAsnTyrAsnLysTyrProGlyProGlnPhe 20
|||||
Db 1 ATCGGAAAAGAAATTCACCTGAATATATACACTATAACAAGTTCTCGTCCGGAGTTC 60
QY 21 IleHisPheGluAsnIleValLysSerAspAspIleGluPheGlnLeuValIleAsnGlu 40
|||
Db 61 CATTTACACGGGACAAAGTCGAAACGGAGGATAGCTTTTCCCTGGTTGAAATATC 120
QY 41 LysSerAlaPheAspValThrValPheGlyGlnArgPheSerGluIleLeuLysTyr 60
|||
Db 121 AAGGATGCTTTCATGTGACGACTTTTATACACGCTTTTCAGAACTATTAAACCAAGTTT 180
QY 61 AspPheIleValGlyAspTrpGlyAsnGluInLeuArgLeuArgGlyPheTyrLysAsp 80
|||||
Db 181 GATTATATCGTGGGGACCTGGAGCAACGACAGCTTCGCCTACGAGGTTTTCACAAAGGAT 240
QY 81 AlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyr 100
|||
Db 241 GACCGACAGAGAAATCTGAAATAATCAGTCGTTTACAGACTACCTTTTAGAGTAT 300
QY 101 CysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspIleLysPhe 120
|||
Db 301 TGTAGTTATGTTGTCCTATTGTTCTTAGAAAATGAAGCCCTTAAGCGAGCATCATTT 360
QY 121 AspAspGluArgProHisLysArgArgLysSerArgSerLysSerGlnSerLysSer 140
|||
Db 361 GAC-----AAGAAATGCGTAGACGGAAGAGAAACACCTTCTAGAAAAGGAGAA 414
QY 141 GlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysLysArg 160
|||
Db 415 CCACCTCAACTAAACGA---AAGTCGATGCGAT----- 447
QY 161 LysAspThrLysArgGlnGluArgHisIleLysGluGluGlnAspLysGluMetThr 180
|||
Db 448 AAGAAAATAGAGCTCGTCAGAAACACGACGATTTCTCAGAAAAGGAGCAAGGAA----- 501
QY 181 SerAlaLysGlnHis-LeuLeuPheValArg 190
|||
Db 502 -----CAACGTCATTTTGTCTATTCGTCGAC 526

RESULT 2

US-09-815-242-9469
;; Sequence 9469, Application US/09815242
;; Patent No. US20020061569A1
;; GENERAL INFORMATION:
;; APPLICANT: Haselbeck, Robert
;; APPLICANT: Ohlsen, Karl L.
;; APPLICANT: Zyskind, Judith W.

;; APPLICANT: Wall, Daniel
;; APPLICANT: Trawick, John D.
;; APPLICANT: Carr, Grant J.
;; APPLICANT: Yamamoto, Robert T.
;; APPLICANT: Xu, H. Howard
;; TITLE OF INVENTION: Identification of Essential Genes in
;; TITLE OF INVENTION: Prokaryotes
;; FILE REFERENCE: ELITRA.011A
;; CURRENT APPLICATION NUMBER: US/09/815,242
;; CURRENT FILING DATE: 2001-03-21
;; PRIOR APPLICATION NUMBER: 60/191,078
;; PRIOR FILING DATE: 2000-03-21
;; PRIOR APPLICATION NUMBER: 60/206,848
;; PRIOR FILING DATE: 2000-05-23
;; PRIOR APPLICATION NUMBER: 60/207,727
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: 60/242,578
;; PRIOR FILING DATE: 2000-10-23
;; PRIOR APPLICATION NUMBER: 60/253,625
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; NUMBER OF SEQ ID NOS: 14110
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 9469
;; LENGTH: 531
;; TYPE: DNA
;; ORGANISM: Streptococcus pneumoniae
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1)...(531)
US-09-815-242-9469

Alignment Scores:
Pred. No.: 2,86e-47 Length: 531
Score: 452.50 Matches: 94
Percent Similarity: 66.49% Conservatives: 33
Best Local Similarity: 49.21% Mismatches: 48
Query Match: 44.85% Indels: 16
DB: 10 Gaps: 4

US-09-868-352-23 (1-192) x US-09-815-242-9469 (1-531)

QY 1 MetArgLysGluValThrProGluMetLeuAsnTyrAsnLysTyrProGlyProGlnPhe 20
|||||
Db 1 ATCGGAAAAGAAATTCACCTGAATATATACACTATAACAAGTTCTCGTCCGGAGTTC 60
QY 21 IleHisPheGluAsnIleValLysSerAspAspIleGluPheGlnLeuValIleAsnGlu 40
|||
Db 61 CATTTACACGGGACAAAGTCGAAACGGAGGATAGCTTTTCCCTGGTTGAAATATC 120
QY 41 LysSerAlaPheAspValThrValPheGlyGlnArgPheSerGluIleLeuLysTyr 60
|||
Db 121 AAGGATGCTTTCATGTGACGACTTTTATACACGCTTTTTCAGAACTTTAAACCAAGTTT 180
QY 61 AspPheIleValGlyAspTrpGlyAsnGluInLeuArgLeuArgGlyPheTyrLysAsp 80
|||||
Db 181 GATTATATCGTGGGGACCTGGAGCAACGACAGCTTCGCCTACGAGGTTTTCACAAAGGAT 240
QY 81 AlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyr 100
|||
Db 241 GACCGACAGAGAAATCTGAAATAATCAGTCGTTTACAGACTACCTTTTAGAGTAT 300
QY 101 CysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspIleLysPhe 120
|||
Db 301 TGTAGTTATGTTGTCCTATTGTTCTTAGAAAATGAAGCCCTTAAGCGAGCATCATTT 360
QY 121 AspAspGluArgProHisLysArgArgLysSerArgSerLysSerGlnSerLysSer 140
|||
Db 361 GAC-----AAGAAATGCGTAGACGGAAGAGAAACACCTTCTAGAAAAGGAGAA 414

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 25, 2003, 11:35:13 ; Search time 67 seconds
(without alignments)
1609.474 Million cell updates/sec

Title: US-09-868-352-23
Perfect score: 1009
Sequence: 1 MRKEVTPMLNKNYKPGPQF.....EEQDKEMTSKQHLLEVRKN 192

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 442118 seqs, 280819700 residues
Total number of hits satisfying chosen parameters: 884236

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09868352/runat_14022003_153518_4305/app_query.fasta_1.391
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=40 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09868352 -CGN_1_1_33 @runat_14022003_153518_4305
-NCFU=6 -ICFU=3 -NO_XLPXY -NO_MMALP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications_NA :
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq : *
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq : *
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq : *
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq : *
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq : *
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq : *
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq : *
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq : *
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq : *
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq : *
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq : *
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq : *
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq : *
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq : *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	452.5	44.8	531	10	US-09-815-242-9316 Sequence 9316, Ap
2	452.5	44.8	531	10	US-09-815-242-9469 Sequence 9469, Ap
3	302.5	30.0	1495	10	US-09-070-927A-502 Sequence 502, App
4	232.5	23.0	285	10	US-09-974-300-7936 Sequence 7936, Ap

5	222.5	22.1	273	10	US-09-974-300-3145 Sequence 3145, Ap
6	99.5	9.9	1534	12	US-10-044-090-268 Sequence 268, App
7	99	9.8	712	10	US-09-867-550-1933 Sequence 1933, Ap
8	99	9.8	3484	10	US-09-925-301-557 Sequence 557, App
9	97.5	9.7	2025	10	US-09-834-975-822 Sequence 822, App
10	96.5	9.6	1428	9	US-10-001-857-79 Sequence 79, Appl
11	95.5	9.5	1343	12	US-10-001-843-70 Sequence 70, Appl
12	95.5	9.5	2212	10	US-09-925-297-142 Sequence 142, Appl
13	95.5	9.5	3259	12	US-10-001-843-71 Sequence 71, Appl
14	93.5	9.3	2269	9	US-09-989-920-113 Sequence 113, App
c	15	92.5	584	10	US-09-878-574-4310 Sequence 4310, Ap
16	92.5	9.2	1267	12	US-10-001-843-45 Sequence 45, Appl
c	17	91.5	1134	10	US-09-864-761-21362 Sequence 21362, A
c	18	91.5	1960	10	US-09-864-761-4620 Sequence 4620, Ap
19	91	9.0	425	10	US-09-960-352-10716 Sequence 10716, A
20	91	9.0	3697	12	US-10-002-600-58 Sequence 58, Appl
21	90.5	9.0	5273	10	US-09-728-952-78 Sequence 78, Appl
22	90.5	9.0	5430	10	US-09-822-635-3 Sequence 3, Appli
23	90.5	9.0	10172	10	US-09-822-635-1 Sequence 1, Appli
c	24	90	272	10	US-09-864-761-18684 Sequence 18684, A
25	90	8.9	1389	9	US-09-938-842A-340 Sequence 340, App
26	89.5	8.9	7521	9	US-09-819-104A-3 Sequence 3, Appli
27	89.5	8.9	8686	9	US-09-819-104A-1 Sequence 1, Appli
c	28	89	203	10	US-09-960-352-5438 Sequence 5438, Ap
c	29	89	489	10	US-09-920-300A-1624 Sequence 1624, Ap
c	30	89	489	12	US-10-033-528-1624 Sequence 1624, Ap
c	31	89	493	10	US-09-920-300A-688 Sequence 688, App
c	32	89	493	12	US-10-033-528-688 Sequence 688, App
c	33	89	609	9	US-09-796-692-4386 Sequence 4386, Ap
c	34	88	170834	10	US-09-835-232-7 Sequence 7, Appli
35	87.5	8.7	333	10	US-09-784-423-27 Sequence 27, Appl
36	87	8.6	575	10	US-09-864-761-20733 Sequence 20733, A
37	87	8.6	1969	10	US-09-864-761-3972 Sequence 3972, Ap
38	87	8.6	2237	9	US-09-832-292-11 Sequence 11, Appl
39	87	8.6	2237	10	US-09-994-485-7 Sequence 7, Appli
c	40	87	8317	10	US-09-764-869-1279 Sequence 1279, Ap
41	86.5	8.6	400	10	US-09-983-965-265 Sequence 265, App
42	86.5	8.6	832	9	US-10-001-857-97 Sequence 97, Appl
43	86.5	8.6	1717	9	US-10-001-857-99 Sequence 99, Appl
44	86.5	8.6	172637	10	US-09-805-458A-3 Sequence 3, Appli
c	45	86	919	10	US-09-878-574-4303 Sequence 4303, Ap

ALIGNMENTS

RESULT 1
US-09-815-242-9316
; Sequence 9316, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27

THIS PAGE BLANK (USPTO)

Search completed: February 25, 2003, 12:00:08
Job time : 1951 secs

Search completed: February 25, 2003, 12:00:08
Job time : 1951 secs


```

Db 233 GTCAACGGCAGAAAATCAGATTAGTTGAAGACAAGCCAGGTTCTAGGCGGCCCGCTCT 292
QY 126 -----HisLysArg-----ArgLysSer 131
Db 293 TACTCCAGAGCGGGAGTCACTCAAGGTCCTCGCTCTGAAGCAGACATTCCTCCGGAAGAGC 352
QY 132 ArgSerLysSerGlnSerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAla 151
Db 353 AGAAGCGGAGTGGTAGCAGTAAAGCAGCCATTCGAAGAGCGGCTCTCGATCCAGGTCA 412
QY 152 AsnAlaHisPheThrSerLysLysArg----- 160
Db 413 GGTTCCTCACTCCGACAGAGAGCCCGCAGTCGCGAGTCAGACGCCGACGCTAGAGAG 472
QY 161 -----LysAspThrLysArgArgGlnGlnArgHisIleLysGlu 173
Db 473 GAGAAAGCGGAGCCCGACCAAGACAACAGAGCCGCGATCGC-----AGCCGC 523
QY 174 GluGlnAspLysGluMetThrSerAlaLysGlnHis 185
Db 524 AGCCCTGACAAGAGCGCGCAGTAGAAGTAAGAGCCAC 559

RESULT 32
BJ111917 652 bp mRNA linear EST 23-JAN-2002
LOCUS BJ111917 unpublished oligo-capped cDNA library, C. elegans L1 stage
DEFINITION Caenorhabditis elegans cDNA clone yk1148h02 5', mRNA sequence.
ACCESSION BJ111917
VERSION BJ111917.1 GI:18271960
SOURCE EST.
ORGANISM Caenorhabditis elegans.
REFERENCE 1 (bases 1 to 652)
AUTHORS Kohara,Y., Shin-i.T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
and Sugano,S.
TITLE A complementary view of the C.elegans genome
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES
Location/Qualifiers
1..652
/organism="Caenorhabditis elegans"
/strain="N2"
/db_xref="taxon:6239"
/clone_lib="yk1148h02"
/elegans_L1_stage="unpublished oligo-capped cDNA library, C. elegans L1 stage"
/sex="hermaphrodite"
/tissue_type="whole animal"
/dev_stage="L1"
/note="The AD-wrmcDNA library was generated with poly(A)+ RNA isolated from both hermaphrodite and male N2 worms of all larval stages, embryos, adults and dauers and the subsequent generation of cDNAs by poly(A) priming. The cDNAs were cloned into pCC86"
BASE COUNT 192 a 135 c 171 g 154 t
ORIGIN

Alignment Scores:
Pred. No.: 0.532 Length: 652
Score: 105.00 Matches: 46
Percent Similarity: 35.53% Conservative: 24
Best Local Similarity: 23.35% Mismatches: 93
Query Match: 10.41% Indels: 34
DB: 13 Gaps: 5

```

```

US-09-868-352-23 (1-192) x BJ111917 (1-652)
QY 11 AsnTyrAsnLysTyrProGlnPheIleHisPheGluAsnIleValLysSerAsp 30
Db 32 AATTACAACAACAACAACATACCATCATCAACAATTTCCGAGAAATCAAT 91
QY 31 AspIle----- 32
Db 92 GAGACCAAGTTCTCTCAGTGAAGCAGACAGATTTTATTTCGAGGACTTCATGGTGATATT 151
QY 33 -----GluPheGlnLeuValIleAsnGlu-----LysSerAlaPheAspVal 46
Db 152 TCGACGGAAGATCAAGGAATATATGGCGAAAGATTGGAAAAATCTCGTTTCGATTTC 211
QY 47 ThrValPheGlyGlnArgPheSerGluIleLeuLeuLysTyrAspPheIleValGlyAsp 66
Db 212 GTGAAAGTGGCTCAAGATAAGAGCAAGATATTCGTGGCTGTTTCGATTTCGAGAATCGTGAT 271
QY 67 TrpGlyAsnGlnLeuArgLeuArgGlyPheTyrLysAspAlaSerThrIleArgLys 86
Db 272 GAGGCTAAGGAGTTTATGGAGACA-----TATTCGATCGTGATTCATGGGATGC 322
QY 87 AsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyrCysAsnPheGlyCysAla 106
Db 323 CGTTGTGATCTCTCTTGGTTCCTCGTAT---ATACGTGCTTATTTGTCCTATCAGCGTGCC 379
QY 107 TyrPheValLeuGluAsnProAsnProArgAspIleLysPheAspAspGluArgProHis 126
Db 380 AAGCAGGTTTCGAAGCAACAGTCAACGCCCGCTCGTCTCGATTTCACAAGAGTCTTCG 439
QY 127 LysArg-----ArgLysSerArgSerLysSerLysSerLysSer 140
Db 440 AAGAGATCTCGTCCGCCACAGTCGAGAGAGGACGATCGAAGAGCGGATCTAGAACGAGA 499
QY 141 GlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysLysArg 160
Db 500 AGTCCCGTCGAGAGCGGATCGAGAGCCGCTCAAGAAAGTGTGTCTAGATCAAGAGTCCCG 559
QY 161 LysAspThrLysArgArgGlnGluArgHisIleLysGluGlnAspLys 177
Db 560 GCATCTAGAAATGGAAGAGTCTCTCGGCATCGTAGTAGAAGTCAGAATCGA 610

RESULT 33
BJ1734757 820 bp mRNA linear EST 20-SEP-2001
LOCUS BJ1734757 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5363670 5',
DEFINITION mRNA sequence.
ACCESSION BJ1734757
VERSION BJ1734757.1 GI:15711770
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 820)
AUTHORS NIH-MGC http://mgi.nhl.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11925 row: d column: 07
High quality sequence stop: 40.
FEATURES
Location/Qualifiers
1..820
/organism="Mus musculus"
/db_xref="taxon:10090"

```

```

Qy 78 TyrLysAspAlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIle 97
Dy 157 TATGCAGATGCTCAACAGGGACGCAAAAC-----GAAGGAGTG 195
Qy 98 LysGluTyrCysAsnPhcGlyCysAlaTyrPheValLeuGlu-----Asn 112
Dy 196 ATTGAATTTGTTCTTACTCTGATATGAAAAGAGCTTTGGAAAAGCTGGACGGAAGTAA 255
Qy 113 ProAsnProArgAspIleLysPheAspGluArgPro----- 125
Dy 256 GTCACGGCAGCAAAATCAGATTAGTTGAAGACAGCCAGGTTCTAGCGCGCGCGTCT 315
Qy 126 -----HisLysArg-----ArgLysSer 131
Dy 316 TACTCCAGAACCGGAGTCACTCAAGGTCGCTCTCGAAGCAGACATTCCTCCGGAAGAGC 375
Qy 132 ArgSerLysSerGlnSerSerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAla 151
Dy 376 AGAAGCCGGAGTGGTAGCAGTAAAGACAGCCATTCGAAGAGCCGGTCTCGATCCAGGTCA 435
Qy 152 AsnAlaHisPheThrSerLysLysArg----- 160
Dy 436 GGTTCCTCCACTCCGAGCAGCAAGCGCGAGTCGAGTCAGAGCCGCGCAGTAAAGAAC 495
Qy 161 -----LysAspThrLysArgArgGlnGluArgHisIleLysGlu 173
Dy 496 GAGAAAGCCGGAGCCCGCAGCAAGACAAACAGAGCCGCGCAGTCGC-----AGCGCG 546
Qy 174 GluGlnAspLysGluMetThrSerAlaLysGlnHis 185
Dy 547 AGCCCTGACAAGAGCCGCGCAGTAAAGAGTAAAGACCAC 582

RESULT 31
BQ569343
LOCUS      BQ569343
DEFINITION gi126h02.y1 Mouse Organ of Corti cDNA pBluescript Mus musculus cDNA
            clone gi126h02 5', mRNA sequence.
ACCESSION BQ569343
VERSION   BQ569343.1 GI:21472660
KEYWORDS  EST.
SOURCE    house mouse.
ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathia; Muridae; Mus.
            Kachar,B.
            1 (bases 1 to 615)
EST analysis of gene expression in the mouse Organ of Corti at the
onset of hearing
Unpublished (2002)
Contact: Kachar,B.
Structural Cell Biology
National Institute of Deafness and other Communication Disorders
50/4249 South Drive, NIH, Bethesda, MD 20892-8027, USA
Tel: 301-402-1599
Fax: 301-402-1765
Email: kacharbenidcd.nih.gov
Plate: 126 row: h column: 02
Seq primer: M13rpl reverse primer (ABI).
            Location/Qualifiers
                1..615
                    /organism="Mus musculus"
                    /strain="BALB/c"
                    /db_xref="taxon:10090"
                    /clone="gil26h02"
                    /clone_lib="Mouse Organ of Corti cDNA pBluescript"
                    /sex="male and female"
                    /dev_stage="Post natal day 5 to 13"
                    /note="Organ of Corti; Vector: pBluescript; The
                    organ of Corti (OC) was fine dissected from a total of 386
                    OC as follows: 102 samples from post-natal (P) day 5; 72
                    from P6; 60 from P7; 46 from P8; 18 from P9; 20 from P10;
                    14 from P12 and 24 from P13. After killing animals by
                    cervical dislocation followed by decapitation, the bulla

```

was removed and opened in Leibowitz medium. The bony capsule of the cochlea was chipped away, stria vascularis and spiral ligament were removed and the sensory epithelium was carefully dissected out of the modiolus. Total RNA was extracted using the micro Fasttrack kit (catalog # K1593-02; Invitrogen, Carlsbad, CA), according to manufacturer's instructions. Reverse transcription and library construction were carried out with the Uni-Zap XR vector kit (catalog # 237211, Stratagene) and Uni-Zap XR GigaPack III Gold Cloning kit (catalog # 237612), both from Stratagene (La Jolla, CA, USA), according to manufacturer's instructions. Briefly: 1.5 ug mRNA was reverse transcribed using a hybrid oligo(dT) linker-primer that contains an Xho I site. First strand synthesis was primed with the linker-primer and transcribed using Moloney murine leukemia virus reverse transcriptase (MMuLV-RT) and 5-methyl dCTP. The second strand was synthesized with DNA polymerase and RNase H. Complementary DNA was blunt ended with Pfu DNA polymerase, ligated with EcoR I adapters in the presence of ligase and digested with Xho I. The cDNA was sequentially size fractionated over Pharmacia Size Sep400 (Pharmacia, Uppsala, Sweden) and Clontech Chroma Spin-1000 (Clontech, Palo Alto, CA) columns to enrich for cDNAs greater than 400bp and 1000 bp, respectively. The cDNA was then directionally ligated to the Uni-Zap XR vector, which had been predigested with EcoR I and Xho I. The phagemid was packaged with GigaPack III Gold and, upon titration on XL1 Blue MRF' cells, the yield of the phage library was estimated to be 11,100,000 recombinants. Stratagene's ExAssist Interference resistance helper phage (catalogue # 211203) was adopted to rescue plasmid DNA from the phages. Upon plating of the rescued library, individual cDNA clones were selected and grown in 96-well, 2 ml growth plate. Plasmid DNA was purified from 200 ul of saturated culture with the Concert96(TM) plasmid purification kit (Invitrogen, Carlsbad, CA) as instructed by the manufacturer. ESTs from the 5' end of the cDNA clones were generated with the universal M13 reverse primer (CAGGAACAGCTATGACC) and 25% strength BigDye terminator sequencing chemistry (Applied Biosystems, Foster City, CA). Sequencing reactions were performed on MJ Tetrad thermal cyclers (MJ Research, Waltham, MA), and analyzed on 3700 automated capillary sequencers using POP5 polymer (Applied Biosystems, Foster City, CA). The frequency distribution of the library is as follows: 72% of genes have 1 copy; 14.3% 2; 12% 3-10; 1.4% 11-50 and 0.1% 51-150. As to gene function, 45% of genes are present in GenBank and have known function; 23% have hits in GenBank, but do not have assigned function; 12% are uncharacterized ESTs and 20% are unidentified."

BASE COUNT 189 a 161 c 176 g 89 t

ORIGIN

Alignment Scores:
 Pred. No.: 0.5 Length: 615
 Score: 105.00 Matches: 40
 Percent Similarity: 36.84% Conservative: 16
 Best Local Similarity: 26.32% Mismatches: 42
 Query Match: 10.41% Indels: 54
 DB: 14 Gaps: 6

US-09-868-352-23 (1-192) x BQ569343 (1-615)

Qy 78 TyrLysAspAlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIle 97
 Dy 134 TATGCAGATGCTCAACAGGGACGCAAAAC-----GAAGGAGTG 172
 Qy 98 LysGluTyrCysAsnPhcGlyCysAlaTyrPheValLeuGlu-----Asn 112
 Dy 173 ATTGAATTTGTTCTTACTCTGATATGAAAAGAGCTTTGGAAAAGCTGGACGGAAGTAA 232
 Qy 113 ProAsnProArgAspIleLysPheAspGluArgPro----- 125


```

SOURCE      human.
ORGANISM     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 731)
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
              Email: cgabbs-r@mail.nih.gov
              Tissue procurement: ATCC
              cDNA Library Preparation: Ling Hong/Rubin Laboratory
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLCM1817 row: h column: 20
              High quality sequence stop: 101.

FEATURES     Location/Qualifiers
              1..731
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:4909507"
                /clone_lib="NIH_MGC_42"
                /tissue_type="epithelioid carcinoma cell line"
                /lab_host="DH10B (phage-resistant)"
                /note="Organ: pancreas; Vector: pOTB7; Site:1: XhoI;
                Site:2: EcoRI; cDNA made by oligo-dT priming.
                Directionally cloned into EcoRI/XhoI sites using the
                following 5' adaptor: GGCACGAG(C). Size-selected >500bp
                for average insert size 1.8kb. Library constructed by Ling
                Hong in the laboratory of Gerald M. Rubin (University of
                California, Berkeley) using ZAP-cDNA synthesis kit
                (Stratagene) and Superscript II RT (Life Technologies).
                Note: this is a NIH_MGC library. |"
              439 a 131 c 125 g 36 t
              BASE COUNT 439 a 131 c 125 g 36 t
              ORIGIN

Alignment Scores:
Pred. No.: 0.535 Length: 731
Score: 105.50 Matches: 34
Percent Similarity: 50.94% Conservative: 20
Best Local Similarity: 32.08% Mismatches: 38
Query Match: 10.46% Indels: 15
DB: 12 Gaps: 4

US-09-868-352-23 (1-192) x BG830845 (1-731)

QY 79 LysAspAlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLys 98
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 262 AAAAAACGAGAACGACAGGACACATAGATCAAGAACACACAAAGAAACAAACAAACAA 321
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 99 GluTyrCysAsnPhgGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspIle 118
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 322 GAGAAAGGAACAAAGGC-----AAAACGGAGCAGCAGCAGCAACCAAGCAAGCAATA 369
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 119 LysPheAspGluArgProHisLys-----ArgArgLysSerArgSerLysSerGln 136
    ::||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 370 -----CAAAAGGAAGCACAAGGAAGACAAAATAAAAAACACAGCAACAAACGACAC 417
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 137 SerSerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPheThr 156
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 418 AAGAAAAACAGACAGCGACACAAAGAAACAAAGAAACAAAGAAACAAAGAAACAA 477
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 157 SerLysLysArg-----LysAspThrLysArgArgGlnGluArgHisIleLys 172
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 478 GAGAAAGAAACGGAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACAA 536
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 173 GluGluGlnAspLysGlu 178
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 537 GAAGACAAAGACAAAGAA 554

```

```

RESULT 30
BG086260
LOCUS
DEFINITION
H3123G04-5 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
H3123G04 5', mRNA sequence.
ACCESSION
BG086260
VERSION
BG086260.1 GI:12568824
KEYWORDS
EST.
SOURCE
house mouse.
ORGANISM
Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 613)
REFERENCE
Kargul,G.J., Dudekula,D.B., Qian,Y., Lim,M.K., Jaradat,S.A., Tanaka
,T.S., Carter,M.G. and Ko,M.S.H.
Verification and initial annotation of NIA mouse 15K cDNA clone set
Unpublished (2001)
JOURNAL
Other_ESTs: H3123G04-3
COMMENT
Contact: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@igsun.grc.nia.nih.gov
This clone set has been freely distributed to the community. Please
visit http://igsun.grc.nia.nih.gov/cDNA/15k.html for details.
Plate: H3123 row: G column: 04
Seq primer: -21M13 Reverse
High quality sequence stop: 613
POLYA=No.

FEATURES     Location/Qualifiers
              1..613
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="niaEST:H3123G04-5"
                /db_xref="taxon:10090"
                /clone="H3123G04"
                /clone_lib="NIA Mouse 15K cDNA Clone Set"
                /sex="Clones arrayed from a variety of cDNA libraries"
                /dev_stage="Clones arrayed from a variety of cDNA
                libraries"
                /lab_host="DH10B"
                /note="Vector: pSPORT1, Site:1: SalI; Site:2: NotI; This
                clone is among a rearranged set of 15,247 clones from 11
                embryo cDNA libraries (including preimplantation stage
                embryos from unfertilized egg to blastocyst, embryonic
                part of E7.5 embryos, extraembryonic part of E7.5 embryos
                , and E12.5 female mesonephros/gonad) and one newborn
                ovary cDNA library. Average insert size 1.5 kb. All
                source libraries are cloned unidirectionally with Oligo(dT
                )-Not primers. References include: (1) Genome-wide
                expression profiling of mid-gestation placenta and embryo
                using a 15,000 mouse developmental cDNA microarray, 2000,
                Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2)
                Large-scale cDNA analysis reveals phased gene expression
                patterns during preimplantation mouse development, 2000,
                Development, 127: 1737-1749; (3) Genome-wide mapping of
                unselected transcripts from extraembryonic tissue of
                7.5-day mouse embryos reveals enrichment in the t-complex
                and under-representation on the X chromosome, 1998, Hum
                Mol Genet 7: 1967-1978."
              191 a 149 c 177 g 96 t
              BASE COUNT 191 a 149 c 177 g 96 t
              ORIGIN

Alignment Scores:
Pred. No.: 0.498 Length: 613
Score: 105.00 Matches: 40
Percent Similarity: 36.84% Conservative: 16
Best Local Similarity: 26.32% Mismatches: 42
Query Match: 10.41% Indels: 54
DB: 12 Gaps: 6

US-09-868-352-23 (1-192) x BG086260 (1-613)

```

COMMENT

Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LAM9767 row: c column: 11
 High quality sequence stop: 566.
 Location/Qualifiers
 1. 1728
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone_lib="IMAGE:4206274"
 /lab_host="NCI_CGAP_Co24"
 /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.6 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library."
 BASE COUNT 599 a 398 c 518 g 213 t
 ORIGIN

FEATURES

source

FEATURES

source

Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LAM10402 row: n column: 01
 High quality sequence start: 4
 High quality sequence stop: 542.
 Location/Qualifiers
 1. 1858
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="IMAGE:4514496"
 /clone_lib="NIH_MGC_93"
 /tissue_type="transitional cell papilloma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: bladder; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dt primed. Average insert size 1.7 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."
 BASE COUNT 703 a 375 c 537 g 243 t
 ORIGIN

Alignment Scores:

Pred. No.: 1.19 Length: 1728
 Score: 106.00 Matches: 27
 Percent Similarity: 44.57% Conservative: 14
 Best Local Similarity: 29.35% Mismatches: 31
 Query Match: 10.51% Indels: 20
 DB: 12 Gaps: 2

US-09-868-352-23 (1-192) x BF578008 (1-1728)

Qy 112 AsnProAsnProArgAspPheAspLysPheAspGluArgProHisLysArgArgLysSer 131
 ||||| ||||| ||| ||||| ||| :|||
 Db 1302 AACCCAGAACGACACACTAAGCAACGACGCGAGAGAGACGCGGAGAACGA 1361
 Qy 132 ArgSerLysSerGlnSerSerLysSerGln----- 141
 |||:||||: ||: |||:||||
 Db 1362 AGAACGAGACGACGCGACAGACACACAAAGAGAGAAAGCGGAACGACACAG 1421
 Qy 142 -----ThrArgAsnAsnArgSerGlnSerAsnAlaAsn 152
 ||||| ||||| ||| |||
 Db 1422 ACAGGAAGACGAGAAACAGACACAAAGACGAGGAGAGAACACACAAACAGAAC 1481
 Qy 153 AlaHisPheThrSerLysLysArgLysAspThrLysArgArgGlnGlu---ArgHisIle 171
 |||: ||| |||:||||| ||: |||
 Db 1482 ACAGCAGCACAGCAACGACGAGAAACAGACAGACGAGCGAAGACGCGTAGGAAGGT 1541
 Qy 172 LysGluGluGlnAspLysGluMetThrSerAlaLys 183
 :||: ||| |||:||||| ||: |||
 Db 1542 GAGAGGAAGCAAGCAAGAGCGCCAGGACCAAGGACCAAGG 1577

RESULT 28

LOCUS BG288695 1858 bp mRNA linear EST 21-FEB-2001
 DEFINITION BG288695F1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4514496 5',
 mRNA sequence.
 VERSION BG288695
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1858)
 REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.

RESULT 29

LOCUS BG830845 731 bp mRNA linear EST 22-MAY-2001
 DEFINITION 602767576F1 NIH_MGC_42 Homo sapiens cDNA clone IMAGE:4909507 5',
 mRNA sequence.
 ACCESSION BG830845
 VERSION BG830845.1 GI:14178432
 KEYWORDS EST.

Alignment Scores:
 Pred. No.: 1.29 Length: 1858
 Score: 106.00 Matches: 30
 Percent Similarity: 42.42% Conservative: 26
 Best Local Similarity: 22.73% Mismatches: 40
 Query Match: 10.51% Indels: 36
 DB: 12 Gaps: 4

US-09-868-352-23 (1-192) x BG288695 (1-1858)

Qy 77 PheTyrLysAspAlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyr 96
 ||||| ||||| ||||| ||||| ||||| |||||
 Db 910 TTCATCTAAGCTGTAAGCACACGAAACGACGACGACGACGACGACGACGACGACG 969
 Qy 97 IleLysGluTyrCysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArg 116
 :||: ||| |||
 Db 970 CAGGAACACAGGTGC-----AGCAAGACACACACAGA 1002
 Qy 117 AspileLysPheAspAspGluArgProHis-----LysArgArgLysSerArgSer 133
 :||: |||||:||||| ||| :|||:|||||
 Db 1003 CGAGCACGACGACGACGACCAAGACACACACTAACGACGACGACGACGACGACGACG 1062
 Qy 134 LysSerGlnSerSerLysSerGln-----ThrArgAsnAsnArgSer 147
 ||||| ||||| ||||| ||||| ||||| |||||
 Db 1063 ATGACCCAGACAGAGAAAGAAAGACGACGACGACGACGACGACGACGACGACGAC 1122
 Qy 148 GlnSerAsnAlaAsnAlaHisPheThrSerLysLys----- 159
 :||| ||| |||||:|||||
 Db 1123 AAGAGCAAGCAGGAATACACAGCAGCAGAGTCGAGAGACGACGACATGCAAGCAGCAGGA 1182
 Qy 160 -----ArgLysAspThrLysArgArgGlnGluArg 169
 |||||:|||||:|||||:|||||:|||||:|||||
 Db 1183 GACGAACACAGACAGAACGCGAGCCATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1242
 Qy 170 HistLysGluGluGlnAspLysGluMetThrSer 181
 :||: ||||| |||||
 Db 1243 CCGAGCAGCAGAGGGAGGACGACCGCAAGACACAGCA 1278

Best Local Similarity: 29.81% Mismatches: 33
 Query Match: 10.51% Indels: 19
 DB: 13 Gaps: 2

US-09-868-352-23 (1-192) x BI552711 (1-878)

Qy 82 SerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyrCys 101
 ||| :|||:|||||:||||| :|||:||||| :|||:||||| :|||:|||||
 Db 622 AGCAACAAAGAAAGAAACACACAGAAACAAAGAAAGACAGACAAAGAAAGAA----- 675
 Qy 102 AsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspIleLysPheAsp 121
 :|||:|||||:||||| :|||:||||| :|||:||||| :|||:|||||
 Db 676 -----AAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 693
 Qy 122 AspGluArgProHisLysArgLysSerArgSerLysSerGlnSerSerLysSerGln 141
 ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 694 AAGAACAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 753
 Qy 142 ThrArgAsnAsnArgSerGlnSerAsnAlaHisPheThrSerLys-LysArgLys 161
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 754 AAAAAACAAATAATAAAAAAAAC 813
 Qy 161 sAspThrLysArgArgGlnGluArgHisIleLysGluGluGlnAspLysGluMetThrSe 181
 ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 814 A-----AAAAGACAGACAGAGAAACGAAACAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 867
 Qy 181 rAlaLysGln 184
 |||:|||||
 Db 868 AAAAAAGAA 877

RESULT 24
 BG490668 1148 bp mRNA linear EST 27-MAR-2001
 LOCUS 602520145F1 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:4638802 5',
 DEFINITION mRNA sequence.

ACCESSION BG490668
 VERSION BG490668.1 GI:13452178

KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1148)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov

Tissue Procurement: DCTD/DTF/Gazdar
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Cloned through the I.M.A.G.E. Consortium/LLNL at:
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LICM401 row: i column: 11
 High quality sequence stop: 150.

FEATURES
 source

1..1148
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4638802"
 /clone_lib="NIH_MGC_18"
 /tissue_type="large cell carcinoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: lung; Vector: pORF7; Site_1: XhoI; Site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned
 into EcoRI/XhoI sites using the following 5' adaptor:
 GGACACGAG(G). Library constructed by Ling Hong in the
 Laboratory of Gerald M. Rubin (University of California,
 Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
 Superscript II RT (Life Technologies). Note: this is a
 NIH_MGC Library."

BASE COUNT 699 a 134 c 289 g 26 t
 ORIGIN

Alignment Scores:
 Pred. NO.: 0.771 Length: 1148
 Score: 106.00 Matches: 28
 Percent Similarity: 41.75% Conservative: 15
 Best Local Similarity: 27.18% Mismatches: 38
 Query Match: 10.51% Indels: 22
 DB: 12 Gaps: 2

US-09-868-352-23 (1-192) x BG490668 (1-1148)

Qy 111 GluAsnProAsnProArgAspIleLysPheAspGluArgProHisLysArgLys 130
 ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 824 GAAACACAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 883
 Qy 131 SerArgSerLysSerGlnSerSerLysSerGlnThrArgAsnAsnArgSerGlnSerAsn 150
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 884 GAAAGGAATAAGGGGAAACAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 940
 Qy 151 AlaAsnAlaHisPheThrSerLysLysArgLys----- 161
 ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 941 CAGAACAAACACGAAACAGAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1000
 Qy 162 -----AspThrLysArgArgGlnGluArg 169
 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 1001 GAGAGAAACACAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1060
 Qy 170 HisIleLysGluGluGlnAspLysGluMetThrSerAlaLysGlnHisLeuPheVal 189
 ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 1061 CACAAAGAGAGAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1120
 Qy 190 ArgLysAsn 192
 |||:|||||
 Db 1121 CGAAAGAAC 1129

RESULT 25
 BF796631

LOCUS BF796631 1260 bp mRNA linear EST 12-JAN-2001
 DEFINITION 602259829F1 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:4343119 5',
 mRNA sequence.

ACCESSION BF796631

VERSION BF796631.1 GI:12101685

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1260)

NIH-MGC http://mgc.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Louis Staudt, M.D., Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Cloned through the I.M.A.G.E. Consortium/LLNL at:
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov

Plate: LLAM9959 row: i column: 08

High quality sequence stop: 313.

Location/Qualifiers

source

1..1260

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4343119"

/clone_lib="NIH_MGC_85"

/tissue_type="lymphoma, cell line"

/lab_host="DH10B (phage-resistant)"

/note="Organ: lymph; Vector: pCMV-SPORT6; Site_1: NotI;

[illegible]

Alignment Scores:

Pred. No.: 0.418 Length: 1256
 Score: 109.00 Matches: 22
 Percent Similarity: 41.55% Conservativeness: 37
 Best Local Similarity: 26.06% Mismatches: 52
 Query Match: 10.80% Indels: 31
 DB: 12 Gaps: 4

US-09-868-352-23 (1-192) x BG121867 (1-1256)

QY 68 GlyAsnGluLeuArgLeuArgGlyPheTyrLysAspAla----- 81
 Db 724 GGAGAGGAGCCCGCAGACGACGCGACGAGCGAGCGCTACAGGAAGGCCCTTGG 783
 QY 82 ---SerThrIleArgLysAsnSerArgIleSer----- 91
 Db 784 GAATCAACCATCTTAGAAGGGAGCAGACACAACCGAGGCAACCATCCGGAGTATCGAG 843
 QY 92 -----ArgLeuGluAspTyrIleLysGluTyrCysAsnPhedGlyCysAla 106
 Db 844 CCCACATAGTCCCTAACATACAGAACAACACAGCTCGTGTCTTCATGACGATA 903
 QY 107 TyrPheValLeuAsnProAsnProArgAspIleLysPheAspGluArgProHis 126
 Db 904 TTGTATCTGCTGCTGCTAACGACGAGACCGCAACATGAGACACACAGTACAG 963
 QY 127 LysArgArgLysSerArgSerLysSerGln-----SerSerLysSerGlnThr 142
 Db 964 AGAAGAAGAGGATCAACAGAAAGAGCAGACGGAATGAGCAGAGCAAGACGAGCAGC 1023
 QY 143 ArgAsn-----AsnArgSerGlnSerAsnAlaAsnAlaHisPheThrSer 157
 Db 1024 GAGACGAAACCCAAAGAAACAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 1083
 QY 158 LysLysArgLysAspThrLysArgArgGlnGluArgHisIleLysGluGluGlnAsp 177
 Db 1084 AGAAGAAGAGAGAACTAAACAGAAACAGACAGACACATAAAGAGAAACAACAAGAG 1143
 QY 177 sglu 178
 Db 1144 AGAA 1147

RESULT 16
 BM880093
 LOCUS ku04h04.y1 Strongyloides ratti PA female naive pAMP1 v1
 DEFINITION Strongyloides ratti cDNA 5', mRNA sequence.
 ACCESSION BM880093
 VERSION
 KEYWORDS EST.
 SOURCE Strongyloides ratti.
 ORGANISM Strongyloides ratti
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Panagrolaimoidea; Strongyloidea; Strongyloides.
 1 (bases 1 to 471)

REFERENCE
 AUTHORS McCarter, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wylie, T., Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagaris, V., Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe, M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.
 TITLE The Washington Univ. Nematode EST Project, 1999
 JOURNAL Unpublished (1999)
 COMMENT Contact: McCarter JP
 The Washington Univ. Nematode EST Project, 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 The library was constructed by Brandi Chiapelli and Dr. James

McCarter at Washington University, St. Louis. The cDNA was made by using Dynabead oligo-dT priming (Dyna). PCR based library using a modified protocol from the SMART PCR cDNA Synthesis Kit from Clontech. Directionally cloned into the UDG sites of pAMP1. Dissected nematode tissues were provided by Dr. Alan Scott. (ascott@jhsph.edu) of the School of Public Hygiene and Public Health at John Hopkins University in Baltimore, MD.
 High quality sequence stop: 395.

FEATURES
 source

1. 471
 Location/Qualifiers
 /organism="Strongyloides ratti"
 /db_xref="taxon:34506"
 /clone_lib="Strongyloides ratti PA female naive pAMP1 v1"
 /dev_stage="parasitic adult females"
 /lab_host="DH10B"
 /note="Vector: pAMP1 (Gibco); Site_1: NotI; Site_2: SalI;
 The library was constructed by Claire Murphy, Brandi Chiapelli and Dr. James McCarter at Washington University, St. Louis. The cDNA was made by using Dynabead oligo-dT priming (Dyna). PCR based library using a modified protocol from the SMART PCR cDNA Synthesis Kit from Clontech. Directionally cloned into the UDG sites of pAMP1. Parasitic adult females were collected from naive animals and provided by Dr. Mark Viney of Bristol, UK."

BASE COUNT 190 a 58 c 108 g 115 t
 ORIGIN

Alignment Scores:
 Pred. No.: 0.208 Length: 471
 Score: 107.50 Matches: 36
 Percent Similarity: 45.64% Conservativeness: 32
 Best Local Similarity: 24.16% Mismatches: 62
 Query Match: 10.65% Indels: 20
 DB: 14 Gaps: 3

US-09-868-352-23 (1-192) x BM880093 (1-471)

QY 46 ValThrValPheGlyClnArgPheSerGluIleLeuLysTyrAspPheIleValGly 65
 Db 52 ATAAATATGATGGGATCATCAATATTGAGTCTTCTCAAAATGCCATATACAGGT 111
 QY 66 -----AspTyrGlyAsnGluGlnLeuArgLeuArg 75
 Db 112 GGTGTTATTATCACTACTATGCAAAATGATGTCAAAATGGCTGCTGAATTTAGT 171
 QY 76 GlyPheTyrLysAspAlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAsp 95
 Db 172 GGTGTACCGAGAGTGGAGAGTAGAAGTGTGAAGATCAAGATCTCTAGTCTTCTTCA 231
 QY 96 TyrIleLysGluTyrCysAsnPhedGlyCysAlaTyrPheValLeuGluAsnProAsnPro 115
 Db 232 TATCTTCTAGTATGGGAGT-----CGAAGAAGTAGATCATCT 270
 QY 116 ArgAspIleLysPheAspGluArgProHisLysArgLysSerArgSerLysSer 135
 Db 271 AGGATAGTAGTAGTATCACTAGAGGATCGAAGAAAGATCATCTAGAAAAAGT 330
 QY 136 GlnSerSerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAlaHisPhe 155
 Db 331 AGG-----AAAAATAGGAAAAGAGATCTAGAGGAGATCTAGATCTAGAAAGTTACTCT 384
 QY 156 ThrSerLysArgLysArgThrLysArgGlnGluArgHisIleLysGluGluGln 175
 Db 385 TCTGATAGAAAAGAAAGATGATCATATAAAAAG-GATAGGAGTGAAGTAGGAGAA 443
 QY 176 AspLysGluMetThrSerAlaLysGln 184
 Db 444 AAAAAGAAAGATAAAGTAGTAAAAAG 470

RESULT 17
 BF672136
 LOCUS
 DEFINITION 602152592F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4293575 5',

1527 bp mRNA linear EST 21-DEC-2000

```

Db 133 ATAGAGTGTGATGTGTGTTGGTGTCTGGTACCGTGAATGAAGAGAGAGAGAG 192
      ::::: ||| ||||| ||| ::::: |||
Qy 78 TyrLysAspAlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIle 97
      ||| ::::: ||| ||| ||| ||| ||| |||
Db 193 TAT-----AGAGAGAGAGAGAGAGAGAGAGCATATATA 222
      :::::
Qy 98 LysGluTyrCysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAsp 117
      ::::: ||| ||| |||
Db 223 TATAAATAT-----ATATAACATGAGTAAATAATGAGAGT 258
      :::::
Qy 118 IleLysPheAspAsp-----GluArgProHisLysArgArgLysSerArgSerLysSer 135
      ||| ::::: ||||| ||| ||| ||| ||| |||
Db 259 GAAAAAGAGAATAATATAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 318
      :::::
Qy 136 GlnSerSerLys-----SerGlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAla 153
      ::::: ||| ||| ||| ||| ||| |||
Db 319 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 378
      :::::
Qy 154 HisPheThrSerLysLysArgLysAspThrLysArgArgGlnGlu---ArgHisIleLys 172
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 379 CACACACACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 438
      :::::
Qy 173 GluGluGlnAspLysGluMetThrSerAlaLys 183
      ||| ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 439 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 471
      :::::
RESULT 14
BF684615 814 bp mRNA linear EST 22-DEC-2000
LOCUS 602140956F1 NIH_MGC_46 Homo sapiens cDNA clone IMAGE:4302205 5',
DEFINITION mRNA sequence.
ACCESSION BF684615.1 GI:11970023
VERSION BF684615.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 814)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LUCM163 row: p column: 14
High quality sequence stop: 382.
Location/Qualifiers
1. .814
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH_MGC_46"
/tissue_type="leiomyosarcoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pOTB7; Site:1: XhoI; Site:2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."
357 a 202 c 154 g 101 t
BASE COUNT
ORIGIN

```

```

Alignment Scores:
Pred. No.: 0.262 Length: 814
Score: 109.00 Matches: 25
Percent Similarity: 52.22% Conservative: 22
Best Local Similarity: 27.78% Mismatches: 33
Query Match: 10.80% Indels: 10
DB: 12 Gaps: 2
US-09-868-352-23 (1-192) x BF684615 (1-814)
Qy 112 AsnProAsnProArgAspIleLysPheAspAspGluArgProHisLysArgArgLysSer 131
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 370 AGCCCAACGACTGACTCTACCTCTACCTCTACCTCTACCTCTACCTCTACCTCTACCTCT 429
      :::::
Qy 132 -----ArgSerLysSerGlnSerSerLysSerGlnThr 142
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 430 AAGACACACGAGACACACACACACACAGTAAAGAAACACGAAACACGAAACACGAAAC 489
      :::::
Qy 143 ArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHis---PheThrSerLysLysArgLys 161
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 490 AAAAAACAAAAGAAAAACACGACGACAGACAGACAGACAGACAGACAGACAGACAGAGAAA 549
      :::::
Qy 162 AspThrLysArgArgGlnGluArgHisIleLysGluGlnAspLysGluMetThrSer 181
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 550 GAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 609
      :::::
Qy 182 AlaLysGlnHisLeuLeuPheValArgLys 191
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 610 AGACAGCAACATATAAAAGAAACTACAGAGAAA 639
      :::::
RESULT 15
BG121867 1256 bp mRNA linear EST 30-JAN-2001
LOCUS 602350882F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:4445541 5',
DEFINITION mRNA sequence.
ACCESSION BG121867
VERSION BG121867.1 GI:12615376
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1256)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10223 row: d column: 22
High quality sequence stop: 571.
Location/Qualifiers
1. .1256
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH_MGC_90"
/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: liver; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.7 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
439 a 303 c 315 g 199 t
BASE COUNT
ORIGIN

```

```

Db 722 CCAACCCCAAGG-----ATGCACAGAAAAAAGAAAAAGAGAGAGA 769
QY 133 SerLysSerGlnSerLysSerGlnThrArgAsnAsnArgSerGln---SerAsnAla 151
Db 770 ACAGAAAGCACAACACAGCAACAGACAGAAAAACAGACACACAAAAAAGCA 829
QY 152 AsnAlaHisPheThrSerLysLysArgLysAspThrLysArgArgGlnGluArgHisIle 171
Db 830 AACACAGACGACACAGCAACAAAGATAAACAGACCAAGGTAAACAAGAAAGAAAAA 889
QY 172 LysGluGluClnAspLysGluMetThrSerAlaLysGln 184
Db 890 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 928

RESULT 12
BG288874 1383 bp mRNA linear EST 21-FEB-2001
LOCUS 602383616F1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4513011 5',
DEFINITION mRNA sequence.
ACCESSION BG288874
VERSION BG288874.1 GI:13044150
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1383)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs@remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10398 row: p column: 04
High quality sequence stop: 146.
FEATURES
Source
1..1383
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4513011"
/clone_lib="NIH_MGC_93"
/tissue_type="transitional cell papilloma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: bladder; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dt primed.
Average insert size 1.7 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 632 a 150 c 439 g 162 t
ORIGIN
Alignment Scores:
Pred. No.: 0.411 Length: 1383
Score: 109.50 Matches: 30
Percent Similarity: 46.67% Conservative: 19
Best Local Similarity: 28.57% Mismatches: 38
Query Match: 10.85% Indels: 18
DB: 12 Gaps: 2
US-09-868-352-23 (1-192) x BG288874 (1-1383)
QY 75 ArgGlyPheTyrLysAspAlaSerThrIleArgLysAsnSerArgIleSerArgLeuGlu 94
Db 560 AGAGGGAGATAGATGCTATATATAGAGAAAAAATTTATATATACAGAGGGGGAG 619
QY 95 AspTyrIleLysGluTyrCysAsnGlyCys-AlaTyrPheValLeuGluAsnProas 114

```

```

Db 620 CTTAT-----TGTTTTATATATAATTATACAGAGAGGAA 655
QY 114 nProArgAspIleLysPheAspGluArgProHisLysArgArgLysSerArgSerLy 134
Db 656 TCAAGG-----ACAGAAAGAGAGAGAGAGAGAGAA 688
QY 134 sSerGlnSerSerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAlaHi 154
Db 689 GAAGACAGAGAAAAAGAAAGGAGAGAACACAGACAGAGAGCGAAACACAGAGAGC 748
QY 154 sPheThrSerLysLysArgLysAspThrLysArgArgGlnGluArgHisIleLysGlu 174
Db 749 GACCCGAGGGAAGAGAGACAGAGAGCAAGAGAAAGACGCGAAGCGGAGAGAGAG 808
QY 174 uGlnAspLysGlu 178
Db 809 GCGAGAGAGAGAG 821

RESULT 13
BG874969 1487 bp mRNA linear EST 20-OCT-2000
LOCUS 601487528F1 NIH_MGC_69 Homo sapiens cDNA clone IMAGE:3889951 5',
DEFINITION mRNA sequence.
ACCESSION BG874969
VERSION BG874969.1 GI:10323745
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1487)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs@remail.nih.gov
Tissue Procurement: DCTD/DTP/Gazdar
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9672 row: g column: 08
High quality sequence start: 56
High quality sequence stop: 75.
FEATURES
source
1..1487
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3889951"
/clone_lib="NIH_MGC_69"
/tissue_type="large cell carcinoma, undifferentiated"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.1 kb. Library constructed by Life
Technologies."
BASE COUNT 933 a 87 c 329 g 138 t
ORIGIN
Alignment Scores:
Pred. No.: 0.445 Length: 1487
Score: 109.50 Matches: 33
Percent Similarity: 43.51% Conservative: 24
Best Local Similarity: 25.19% Mismatches: 51
Query Match: 10.85% Indels: 23
DB: 12 Gaps: 5
US-09-868-352-23 (1-192) x BG874969 (1-1487)
QY 58 LeuLysTyrAspPheIleValGlyAspTrpGlyAsnGluClnLeuArgLeuArgGlyPhe 77

```

```

Oy 153 AlaHisPheThrSerLysLysArgLysAspThrLysArgGlnGlnLysHisLysLeu 172
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 593 AAA-----ACAGAGAAAAAATAAGACACAAAAACAGACACAAAAAGAG 646
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 173 GluGluGlnAspLysGluMet-----ThrSerAla 182
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 647 CAAGAAGACGTAAAGACATCAGAACCAAGAAAAAAGACCAAGCAAGCAACACA 706
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 183 LysGlnHisLeuLeuPheValArgLysAsn 192
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 707 AAAAACCAACACATAAAAGACGACGAAC 736
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
LOCUS AL640539 505 bp mRNA linear EST 12-DEC-2001
DEFINITION AL640539 XGC-neurula Silurana tropicalis cDNA clone TNeu04k13 5',
    mRNA sequence.
ACCESSION AL640539
VERSION AL640539.1 GI:16792670
KEYWORDS EST.
SOURCE western clawed frog.
ORGANISM Silurana tropicalis
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipelidae;
    xenopodinae; Silurana.
REFERENCE 1 (bases 1 to 505)
AUTHORS Huckle,E., Taylor,R., Ashurst,J.L., Zorn,A.M. and Rogers,J.
TITLE Sanger Xenopus tropicalis EST project 2001 (10-2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Huckle E
    Sanger Centre
    Hinxton, Cambridgeshire, CB10 1SA, UK
    Email: tropesanger.ac.uk
    Sanger Xenopus tropicalis EST project 2001
    TROPICALIS_SEQUENCE_ID: TNeu04k13.sp6
    Sequencing primer: SP6
    This sequence is from a Xenopus Gene Collection (XGC) library
    constructed by Aaron M. Zorn.
FEATURES
    source
        1..505
            location/Qualifiers
                /organism="Silurana tropicalis"
                /db_xref="taxon:8364"
                /clone="TNeu04k13"
                /clone_lib="XGC-neurula"
                /dev_stage="neurula"
                /lab_host="Escherichia coli DH10B"
                /note="vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
                was oligo dt primed from Sug of poly A+ RNA from neurula.
                EcoRI-NotI cut cDNA was then ligated into pCS107 with
                EcoRI at the 5' end and NotI at the 3' end."
BASE COUNT 105 a 178 c 145 g 76 t 1 others
ORIGIN
Alignment Scores:
    Pred. No.: 0.14 Length: 505
    Score: 109.50 Matches: 36
    Percent Similarity: 46.03% Conservative: 22
    Best Local Similarity: 28.57% Mismatches: 56
    Query Match: 10.85% Indels: 12
    DB: 9 Gaps: 3

US-09-868-352-23 (1-192) x AL640539 (1-505)
Oy 68 GlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAspAlaSerThrIleArgLysAsn 87
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 34 GGAGAGCGGGGCTTCGCTTCGTTCACACAAAGCGCGATGCCGAAGACGCGAT 93
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 88 SerArg-IleSer-----ArgLeuGluAspTyrIleLysGluTy 100
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 94 GGACGCCATGATGGGCGCGTCTCGACGGCGGAGCTGAGGTGCAGATGGCCCGGTA 153
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 100 rCysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArg-----As 117

```

```

|||||: :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 154 CTGTTCTTCTCCCGATTCCCAACCAAGGAGAGGGCCCTCTCTCCGACATACGGGA 213
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 117 pileLysPheAspAspGluArgProHisLysArgArgLysSerArgSerLysSerGlnSe 137
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 214 CTAGCGCGCGGAGCAGAGAGAGCCAGGAGCGACGTCGCGACGCGTTCCAGAACAGAG 273
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 137 r---SerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPheTh 156
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 274 CAGTCAAGGTCCCGCAGCAGGTCTCGCTACAGCCGCTCAAGTCCCGCTCTCCGACACAG 333
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 156 rSerLysLysArgLysAspThrLysArgGlnGlnArgHisLysLysGluGluGlnAs 176
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 334 GTCCCGCAGCCGCTCCAGTCAAAGTCTCGCTCGCAAGGAGATCCCAAGTCTAAGTCATC 393
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 176 pLysGluMetThrSer 181
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 394 GTCCGCGCTCCAGGTCT 409
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
LOCUS BG259084 1005 bp mRNA linear EST 13-FEB-2001
DEFINITION 602379137F1 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:4509800 5',
    mRNA sequence.
ACCESSION BG259084
VERSION BG259084.1 GI:12768887
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1005)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
    Email: cgapbs-r@mail.nih.gov
    Tissue Procurement: ATCC
    CDNA Library Preparation: Life Technologies, Inc.
    CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
    DNA Sequencing by: Incyte Genomics, Inc.
    Clone distribution: MGC clone distribution information can be
    found through the I.M.A.G.E. Consortium/LLNL at:
    http://image.llnl.gov
    Plate: L1AM10390 row: j column: 09
    High quality sequence stop: 438.
FEATURES
    source
        1..1005
            location/Qualifiers
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:4509800"
                /clone_lib="NIH_MGC_92"
                /tissue_type="embryonal carcinoma, cell line"
                /lab_host="DH10B (phage-resistant)"
                /note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
                Site_2: SalI; Cloned unidirectionally; oligo-dt primed.
                Average insert size 2.5 kb. Library enriched for
                full-length clones and constructed by Life Technologies.
                Note: this is a NIH_MGC Library."
BASE COUNT 467 a 202 c 211 g 125 t
ORIGIN
Alignment Scores:
    Pred. No.: 0.292 Length: 1005
    Score: 109.50 Matches: 25
    Percent Similarity: 56.16% Conservative: 16
    Best Local Similarity: 34.25% Mismatches: 27
    Query Match: 10.85% Indels: 5
    DB: 12 Gaps: 2

US-09-868-352-23 (1-192) x BG259084 (1-1005)
Oy 113 ProAsnProArgAspIleLysPheAspGluArgProHisLysArgArgLysSerArg 132

```

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="niaest:L0230G12-3"
/db_xref="taxon:10090"
/clone="L0230G12"
/clone_lib="NIA Mouse Newborn Ovary cDNA Library"
/sex="female"
/dev_stage="Newborn Ovary"
/lab_host="DH10B"
/notes="Vector: pSPORT1 (Gibco/BRL Life Technology);
Site 1: SalI; Site 2: NotI; Total RNAs were extracted from
7 Newborn Ovary. The double-stranded cDNA was synthesized
by Gibco's kit with an oligo(dT) primer [NotI
primer-adaptor from Gibco/BRL]
[5'-GGACTAGTCTAGATCGGAGCGGCCCTTTT-3'] from
2.5ug of total RNA. The double-stranded cDNAs were
treated with T4 DNA polymerase and purified by
ethanol-precipitation. The cDNAs were ligated to
lone-linker LL-Sal3 (include SalI sequence). The cDNAs
were purified by phenol/chloroform and separated from
free linkers by Centricon 100. Then, cDNAs were amplified
by long-range high fidelity PCR using Takara's Ex Taq
polymerase. Then, the cDNAs were purified by
phenol/chloroform and by Centricon 100. The cDNAs were
digested with SalI and NotI enzymes. Then, the cDNAs were
size selected by Gibco's Size Fractionation Column. The
cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid
vector. The DH10B E. coli host was transformed with the
ligation mixture by chemical method. The library was
constructed by Xiaohong Wang and Yulan Piao."
BASE COUNT      114 a 154 c 102 g 181 t
ORIGIN
Alignment Scores:
Pred. No.:      0.121      Length:      551
Score:          110.50      Matches:      30
Percent Similarity: 49.44%      Conservative: 14
Best Local Similarity: 33.71%      Mismatches: 34
Query Match:      10.95%      Indels:      11
DB:              10      Gaps:      3
US-09-868-352-23 (1-192) x AW553719 (1-551)
QY 103 PheGlyCysAla-----TyrPheValLeuGluAsnProAsnPro 115
Db 384 TTTAATTGTTCTGAGCTTACGCTGCAAGTGGTTTTTCTCCAGAGTCTGGAAAGAGC 325
QY 116 ArgAspIleLysPheAspAspGluArgProHisLysArgArgLysSerArgSerLysSer 135
Db 324 AATGAAGGAAGGCGGTGATCTGTTCCACACTCGCTCAAGTCCAGGTCT----- 271
QY 136 GlnSerSerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPhe 155
Db 270 ---AGTCAAAATCCCACCTCTAGAAAGAGAGATCATCAGTCAAGACACAGATCCCATAT 214
QY 156 ThrSerLysLysArg---LysAspThrLysArgArgGlnGluArgHisIleLysGluGlu 174
Db 213 AGTCTACGCTCAAGACAGAAGATAGATAGATCCAAAGAGCCACACACAAAGACGCTCT 154
QY 175 GlnAspLysGluMetThrSerAlaLys 183
Db 153 AAGTCTAGGAGAGGCGGAAGTCAAGG 127

```

```

RESULT 9
BE899526
LOCUS      BE899526      999 bp      mRNA      linear      EST 29-SEP-2000
DEFINITION      601681586F1 NIH_MGC_9 Homo-sapiens cDNA clone IMAGE:3951885 5',
                mRNA sequence.
ACCESSION      BE899526
VERSION      BE899526.1      GI:10367116
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens

```

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: DCTD/DTF
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at: Image.lnl.gov
Plate: LNCM820 row: k column: 22
High quality sequence start: 5
High quality sequence stop: 361.
FEATURES
Location/Qualifiers
1..999
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3951885"
/clone_lib="NIH_MGC_9"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

```

BASE COUNT      459 a 203 c 253 g 83 t 1 others
ORIGIN
Alignment Scores:
Pred. No.:      0.229      Length:      999
Score:          110.50      Matches:      47
Percent Similarity: 38.42%      Conservative: 26
Best Local Similarity: 24.74%      Mismatches: 66
Query Match:      10.95%      Indels:      51
DB:              12      Gaps:      7
US-09-868-352-23 (1-192) x BE899526 (1-999)
QY 31 AspIleGluPheGlnLeuValIleAsnGluLysSerAlaPheAspValThrValPheGly 50
Db 236 GATGCTGAGGAAGATTGTGTGGTCCAGCCAGCAGTGGAGATCAACAGGAGGGA 295
QY 51 GlnArgPhe-----SerGluIleLeuLeuLysTyr 60
Db 296 GACACTTCTTACATCAAAACCTCCACCAGTGGCCACACAGAGATT----- 343
QY 61 AspPheIleValGlyAspTrpGlyAsnGluGlnLeuArgLeu-----ArgGlyPhe 77
Db 344 AACTTCAAGTGGGAGGAGTAGAGAGACAGACAGTGGATGGAGAGCGCAAAAGCTGG 403
QY 78 TyrLysAspAlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIle 97
Db 404 GCAAGGGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 448
QY 98 LysGluTyrCysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAsp 117
Db 449 -----GGAGAAAGGCCCAAGAAACAGGGA 472
QY 118 IleLys-----PheAspAspGluArgProHisLysArgArgLysSerArg 132
Db 473 GCAAAAGAACAGAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 532
QY 133 SerLysSerGlnSerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAlaAsn 152
Db 533 GACTACAGCACACAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 592

```


Percent Similarity:	49.49%	Conservative:	26
Best Local Similarity:	23.23%	Mismatches:	45
Query Match:	11.05%	Indels:	5
DB:	9	Gaps:	2

US-09-868-352-23 (1-192) x AI404883 (1-471)

Qy	91	SerArgLeuGluAspTyrIle---	LysGluTyrCysAsnPhcClyCysAlatYrPheVal	109
		::::		
Db	34	TCTGTATAATGATTATTTT	GTAAGCGTATGTAATTTTCACACTATTTTCC	93
Qy	110	LeuGluAsnProAsnProArgaspIleLysPheAspGluArgProHisLysArgArg	129	
			::: :::::	
Db	94	CGTCCAATGCAATAAATAAAATGTTAACTAGAAAAAAA	AAAAAAAAAAAAAAAAAAAA	153
Qy	130	LysSerArgSerLysSerGlnSerSerLysSerGlnThrArgAsnAsnArgSerGlnSer	149	
		::::	:	
Db	154	AAAAAAAAAAAAAAAAAAAA	AAAAAAAAAAAAAAAAAAAA	213
Qy	150	AsnAlaAsnAlaHisPheThrSerLysLysArgGluArgThrLysArgArgGlnGluArg	169	
		::::		
Db	214	AAAAAAAAA-----	AAAAAAAAAAAAAAAAAAAA	261
Qy	170	HisLysGluGluGlnAspLysGluMetThrSerAlaLysGlnHisLeuLeuPhe	188	
		:::::		
Db	262	AAAAAAAAAAAAAAAAAAAA	AAAAAAAAAAAAAAAAAAAA	318

RESULT 5	
BI082113	
LOCUS	951 bp mRNA linear EST 20-JUN-2001
DEFINITION	602877276f1 NCI_CGAP_Mam2 Mus musculus cDNA clone IMAGE:5008939 5',
ACCESSION	BI082113
VERSION	BI082113.1
KEYWORDS	GI:14500443
SOURCE	EST.
ORGANISM	house mouse Mus musculus
REFERENCE	Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 951)
TITLE	NIH-MGC http://mgc.nci.nih.gov/ .
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgabbs@email.nih.gov Tissue procurement: Gilbert Smith, Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM1054 row: g column: 20 High quality sequence stop: 420.

Score:	111.50	Matches:	27
Percent Similarity:	61.54%	Conservative:	21
Best Local Similarity:	34.62%	Mismatches:	19
Query Match:	11.05%	Indels:	11
DB:	13	Gaps:	3

US-09-868-352-23 (1-192) x BI082113 (1-951)

Qy	116	ArgAspIleLysPheAspAsp-----GluArgProHis-----	126
Db	471	CGACACCACAAGAGAGAGGACAGAAGAAAGAACACACAGGGGAAGAGAGCGCACGAGA	530
Qy	127	LysArgArgLysSerArgSerLysSerGlnSerSerLysSerGlnThrArgAsnAsnArg	146
Db	531	AAAGAGCGCAGAAACAACACAGACACACGACGCAACGAAAACGACGACCAAGACACGCG	590
Qy	147	SerGlnSerAsnAlaAsnAlaHisPhetHrSerLysLysArgLysAspThrLysArgArg	166
Db	591	ACACTGAGCCAGACCCACAGGCACACGACACACAANGAAGAGAGAGAAACACAAACAGACAG	650
Qy	167	GlnGluArgHisLleLysGluGluGlnAspLysMetThrSerAlaLysGln	184
Db	651	AAAGAAAAG-----AAGGAGGAAAGAAACACAGACACGCCCGCCCAAGAGAACAA	698

linear EST 03-APR-2001
676 bp mRNA Homo sapiens cDNA clone IMAGE:4662805'

RESULT 6	BG534042
LOCUS	602553149F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4662805', mRNA sequence.
ACCESSION	BG534042
VERSION	BG534042.1 GI:13525582
KEYWORDS	EST.
SOURCE	human. ; human.
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 676) NIH-MGC http://mgc.nci.nih.gov/ . National Institutes of Health, Mammalian Gene Collection (MGC). Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs@mail.nih.gov
AUTHORS	Tissue Procurement: CLONTECH Laboratories, Inc. CDNA Library Preparation: CLONTECH Laboratories, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
JOURNAL	DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
COMMENT	Plate: LICMI464 row: a column: 14 High quality sequence stop: 231.

Location/Vocallibels

1. .676

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:462805"

/clone_lib="NIH_MGC-77"

/lab_host="DH10B (T1 phage-resistant)"

/note="Organ: lung; Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggccattcgccc); Site_2: SfiI (ggccattcgccc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CAGCGCCATTATGCC-3' and 3' adaptor sequence:
5'-ATTCTAGGCGAGGCGGGCGGCATG-dT(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.9
kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH_MGC Library."

336 a 133 c 145 g 62 t

BASE COUNT
ORIGIN

```

5', mRNA sequence.
B0887742
VERSION B0887742.1 GI:22279756
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
TITLE NIH-MGC http://mgi.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue procurement: Susan L. Sullivan, PhD.
CDNA Library Preparation: ResGen, Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LAM13744 row: i column: 12
High quality sequence stop: 656.
FEATURES
    source
        1..951
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /clone="IMAGE:6314963"
            /clone_lib="NIH-MGC_129"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: olfactory epithelium; Vector:
            PCMV-SPORT6.1.cdb; Site.1: EcorV; Site.2: NotI; Cloned
            unidirectionally. Primer: Oligo dr. Average insert size
            2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this
            is a NIH-MGC Library."
BASE COUNT 284 a 231 c 293 g 141 t 2 others
ORIGIN
Alignment Scores:
Pred. No.: 0.107 Length: 951
Score: 113.50 Matches: 58
Percent Similarity: 35.97% Conservative: 33
Best Local Similarity: 22.92% Mismatches: 71
Query Match: 11.25% Indels: 91
DB: 14 Gaps: 12
US-09-868-352-23 (1-192) x B0887742 (1-951)
Qy 14 LysTyrProGlyProGln-----PheIleHisPheGluAsnIleValLysSerAspAsp 31
Db 6 GAATTCGGGATCAAGAGATGGCTTTGTGGAGTTTGATGATCGTGGATGCAGACCAT 65
Qy 32 IleGluPheGlnLeuValIleAsnGluLysSerAlaPheAspValThrVal----- 48
Db 66 GCTGTTTATGAACATGACGCGCAAGACCTGTGTGCGAGCGAGTCATGTTGAGCACGCT 125
Qy 49 -----PheGlyClnArgPhe 53
Db 126 CGAGCGCGCGGGATGGCAGCTACCGCTCGGACCGCAGTGGATGGTATCGAAGA 185
Qy 54 Ser-----GluIleLeuLeuLysTyrAspPheIleValGlyAsp 66
Db 186 AGTGGCGGAGATAATACGGTCCCTACTCGCAGAGATACAGACTTATTGTGGAGAT 245
Qy 67 -----TrpGlyAsn-----GluGlnLeuArgLeuArgGly----- 76
Db 246 TTGTCAAGTCGATGCGTGGCAGGACCTAAAGGATTACATGCGTCAGGCGAGGAGTG 305
Qy 77 PheTyrLysAspAlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyr 96
Db 306 ACGTATGAGATGCTACACAGGAGCGCAAAAC-----GAAGGA 344
Qy 97 IleLysGluTyrCysAsnPheGlyCysAlaTyrPheValLeuGlu----- 111

```

```

... |||... :... |||...
Db 345 GTGATTGAATTGTCTCTTACTCTGATGAAAAGAGCTTTGAAAAGCTGGACGGAAC 404
Qy 112 AsnProAsnProArgAspIleLysPheAspAspGluArgPro----- 125
Db 405 GAAGTCAACGGCAGAAAATCAGATTAGTTGAAGACAAGCCAGGTTCTAGCGCGCGCGG 464
Qy 126 -----HisLysArg-----ArgLys 130
Db 465 TCTTACTCCAGAGCGCGAGTCACTCAAGGTCCTCTCGAAGCAGACATTCGCGGAG 524
Qy 131 SerArgSerLysSerGlnSerLysSerLysSerGlnThrArgAsnAsnArgSerGlnSerAsn 150
Db 525 AGCAGAAGCGGAGTGGTGCAGTAAAGACGAGCCATTTCGAAGAGCGCGTCTCGATCCAGG 584
Qy 151 AlaAsnAlaHisPheThrSerLysLysArg----- 160
Db 585 TCAGGTTCCCACTCCCGCAGCAAGAGCGCGAGTCCAGTCAGAGCGCGCAGCAGTAAG 644
Qy 161 -----LysAspThrLysArgArgGlnGlnLysLysLys 172
Db 645 AAGGAGAAAGCGCGAGCGCGCAGCAACACACAGAGCGCGCAGTCGC-----AGC 695
Qy 173 GluGluGlnAspLysGluMetThrSerAlaLysGlnHis 185
Db 696 CGCAGCCCTGACAGAGCGCGCAGTAAAGAGTAAGACCAAC 734
RESULT 4
LOCUS A1404883 471 bp mRNA linear EST 23-APR-2001
DEFINITION GH24733.5prime GH Drosophila melanogaster head pOT2 Drosophila
melanogaster cDNA clone GH24733 5, mRNA sequence.
ACCESSION A1404883
VERSION A1404883.1 GI:4247970
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
REFERENCE Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
AUTHORS Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
TITLE Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,
JOURNAL BDCP/HMI Drosophila EST Project
COMMENT Unpublished (2001)
Other_ESTs: GH24733.3prime
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: GH.247 row: C column: 9
High quality sequence stop: 322
POLYA-No.
FEATURES
    source
        1..471
            /organism="Drosophila melanogaster"
            /db_xref="taxon:7227"
            /clone="GH24733"
            /clone_lib="GH Drosophila melanogaster head pOT2"
            /sex="male and female"
            /dev_stage="adult"
            /lab_host="DH5 - alpha"
            /note="Organ: head; Vector: pOT2; Site.1: EcoRI; Site.2:
            XhoI; Sized fractionated cDNAs were directly ligated into
            pOT2. Plasmid cDNA library."
BASE COUNT 289 a 36 c 27 g 119 t
ORIGIN
Alignment Scores:
Pred. No.: 0.0809 Length: 471
Score: 111.50 Matches: 23

```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 25, 2003, 10:34:18 : Search time 1942 Seconds
(without alignments)
1601.202 Million cell updates/sec

Title: US-09-868-352-23
Perfect score: 1009
Sequence: 1 MRKEVTPMLNKNYPGPQF.....EEODKEMTSAKQHLFLVRKN 192

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09868352/runat_14022003_153516_4230/app_query.fasta_1.391
-DB=EST -QPMT=fastap -SUFFIX=1st -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=40 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09868352 @CGN_1.1.1349 @runat_14022003_153516_4230 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-Ygapop=10 -Ygapext=0.5 -DELOP=6 -DELEXT=7

Database :

EST: *
1: em_estba: *
2: em_esthm: *
3: em_estin: *
4: em_estnu: *
5: em_estov: *
6: em_estpl: *
7: em_estro: *
8: em_htc: *
9: gb_est1: *
10: gb_est2: *
11: gb_htc: *
12: gb_est3: *
13: gb_est4: *
14: gb_est5: *
15: em_estfun: *
16: em_estom: *
17: gb_gss: *
18: em_gss_hum: *
19: em_gss_inv: *
20: em_gss_pin: *
21: em_gss_vrt: *
22: em_gss_fun: *
23: em_gss_mam: *
24: em_gss_mus: *
25: em_gss_other: *
26: em_gss_pro: *
27: em_gss_rod: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	200.5	19.9	4933	17	BH771017	BH771017 LLMGtag73
2	113.5	11.2	686	12	BE879112	BE879112 601485839
3	113.5	11.2	951	14	BQ887742	BQ887742 AGENCOURT
4	111.5	11.1	471	9	AI404883	AI404883 GH24733.5
5	111.5	11.1	951	13	BI082113	BI082113 60287276
6	111	11.0	676	12	BG534042	BG534042 602553149
7	111	11.0	1089	12	BG850731	BG850731 102402981
c	110.5	11.0	551	10	AW553719	AW553719 L0230612
9	110.5	11.0	999	12	BE899526	BE899526 601681586
10	109.5	10.9	505	9	AL640539	AL640539 AL640539
11	109.5	10.9	1005	12	BG259084	BG259084 602379137
12	109.5	10.9	1383	12	BG288874	BG288874 60233816
13	109.5	10.9	1487	12	BE874969	BE874969 601487528
14	109	10.8	814	12	BF684615	BF684615 602140956
15	109	10.8	1256	12	BG121867	BG121867 602350882
16	107.5	10.7	471	14	BM880093	BM880093 ku04h04.v
17	107.5	10.7	1527	12	BF672136	BF672136 602152592
18	107	10.6	1091	13	BG971957	BG971957 602841742
19	106.5	10.6	511	9	AA838525	AA838525 oes90h12.s
20	106.5	10.6	966	10	BB619189	BB619189 BB619189
21	106.5	10.6	1932	12	BG114689	BG114689 602315330
22	106	10.5	831	13	BG922949	BG922949 602823807
23	106	10.5	878	13	BI552711	BI552711 603193962
24	106	10.5	1148	12	BG490668	BG490668 602520145
25	106	10.5	1260	12	BF796631	BF796631 602239829
26	106	10.5	1572	12	BG285821	BG285821 602389913
27	106	10.5	1728	12	BF578008	BF578008 602091842
28	106	10.5	1858	12	BG288695	BG288695 602385507
29	105.5	10.5	731	12	BG830845	BG830845 602767576
30	105	10.4	613	12	BG086260	BG086260 h3123504
31	105	10.4	615	14	BQ569343	BQ569343 g1126h02
32	105	10.4	820	13	BJ111917	BJ111917 BJ111917
33	105	10.4	978	12	BE914647	BE914647 601665155
34	105	10.4	1122	12	BG468632	BG468632 602510232
35	105	10.4	1173	12	BG519043	BG519043 602577966
36	105	10.4	1267	12	BE839084	BE839084 RC6-EN013
c	104.5	10.4	388	12	BE839084	BE839084 RC6-EN013
39	104.5	10.4	622	12	BG885312	BG885312 dad40c05
40	104.5	10.4	1166	11	AK007729	AK007729 Mus muscu
41	104.5	10.4	1787	12	BG256691	BG256691 602370880
42	104	10.3	882	12	BF213994	BF213994 601847482
43	104	10.3	911	12	BG478911	BG478911 602525126
44	104	10.3	929	12	BG529461	BG529461 602557839
45	104	10.3	1197	13	BI733626	BI733626 603354814

ALIGNMENTS

RESULT 1	BH771017	4933 bp	DNA	linear	GSS 01-MAY-2002
LOCUS	LLMGtag739	MG1363	Random Sequence Tag	Library	Lactococcus lactis
DEFINITION	subsp. cremoris genomic, DNA sequence.				
ACCESSION	BH771017				
VERSION	BH771017.1	GI:20373974			
KEYWORDS	GSS.				
SOURCE	Lactococcus lactis subsp. cremoris.				
ORGANISM	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.				
REFERENCE	1 (bases 1 to 4933)				
AUTHORS	Bolotin.A., Ehrlich.S.D. and Sorokin.A.				
TITLE	Studies of genomes of dairy bacteria Lactococcus lactis				

THIS PAGE BLANK (USPTO)

```
QY 87 AsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyrCys----- 101
|||... ||| :... |||
Db 91958 AACAAATCAACCCAGGAGGAGGTCACACATGTGTGTGTACCTCACAAG 92017
|||... ||| :... |||

QY 102 -----AsnPheClyCysAlaTyrPheValLeuGluAsnProAsnProArgAsp 117
|||... ||| :... |||
Db 92018 TTACAAATCTGGAACTTTGGACACAGATGGCTTATGTGTGAAAGTCCAAAGTCAAGAAGT 92077
|||... ||| :... |||

QY 118 IleLysPhe----- 120
|||... ||| :... |||

Db 92078 GGGAGGATGAGCACAAGAGAAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 92137
|||... ||| :... |||

QY 121 ---AspAspGluArgProHisLysArgLysSerArgLysSerLysSerGlnSerSerLys 139
|||... ||| :... |||
Db 92138 AAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 92197
|||... ||| :... |||

QY 140 SerGlnThrArgAsnAsnArgSerGlnSerAsnAlaHisPheThrSerLysLys 159
|||... ||| :... |||
Db 92198 AAGAAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 92257
|||... ||| :... |||

QY 160 ArgLysAspThrLysArgArgGlnClnArgHisIleLysGluGlnAspLysGluMet 179
|||... ||| :... |||
Db 92258 AGGAAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 92317
|||... ||| :... |||

QY 180 ThrSerAlaLysGln 184
|||... ||| :... |||
Db 92318 ACGAGCAGAGAGAAA 92332
|||... ||| :... |||
```

Search completed: February 25, 2003, 11:35:08
Job time : 3159 secs

ORGANISM	REFERENCE	AUTHORS
Rattus norvegicus	1 (bases 1 to 168111)	Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C., Alsbrooks, S.L., Amaratunge, H.C., Are, J.R., Ayale, M., Banks, T., Barbara, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseghe, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, N., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenwo, S., Oguh, M., Okwuonu, G., Oragunye, N., Ovierson, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Roife, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalob, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K., Wu, C., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G. and Gibbs, R.
Rattus norvegicus	2 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	3 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	4 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	5 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	6 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	7 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	8 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	9 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	10 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	11 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	12 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	13 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	14 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	15 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	16 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	17 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	18 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	19 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	20 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	21 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	22 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	23 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	24 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	25 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	26 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	27 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	28 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	29 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	30 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	31 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	32 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	33 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	34 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	35 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	36 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	37 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	38 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	39 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	40 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	41 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	42 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	43 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	44 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	45 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	46 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	47 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	48 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	49 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	50 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	51 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	52 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	53 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	54 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	55 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	56 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	57 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	58 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	59 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	60 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	61 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	62 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	63 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	64 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	65 (bases 1 to 168111)	Worley, K.C.
Rattus norvegicus	66 (bases 1	

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE AUTHORS

1 (bases 1 to 132934)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Anaratunge,H.C., Are,J.R., Avele,M., Banks,T., Barbaria,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brivea,M., Brown,E., Brown,M., Bryant,N.P., Buhatz,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delantey,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoshitari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G., and Gibbs,R.

Direct Submission
Unpublished
2 (bases 1 to 132934)
Worley,K.C.

Direct Submission
Submitted (03-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 132934)
Worley,K.C.

Direct Submission
Submitted (18-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jul 14, 2002 this sequence version replaced gi:20799942.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GVM1
Center clone name: CH230-357A23
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 105131 bases at least Q40
Consensus quality: 108098 bases at least Q30

Consensus quality: 109837 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 40 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1
* 1186: contig of 1186 bp in length
* 1187
* 1286: gap of unknown length
* 1287
* 2398: contig of 1112 bp in length
* 2399
* 2498: gap of unknown length
* 2499
* 3572: contig of 1074 bp in length
* 3573
* 3672: gap of unknown length
* 3673
* 4761: contig of 1089 bp in length
* 4762
* 4861: gap of unknown length
* 4862
* 6408: contig of 1547 bp in length
* 6409
* 6508: gap of unknown length
* 6509
* 8289: contig of 1781 bp in length
* 8290
* 8389: gap of unknown length
* 8390
* 9557: contig of 1168 bp in length
* 9558
* 9657: gap of unknown length
* 9658
* 10777: contig of 1120 bp in length
* 10778
* 10877: gap of unknown length
* 10878
* 12603: contig of 1726 bp in length
* 12604
* 12703: gap of unknown length
* 12704
* 14881: contig of 2178 bp in length
* 14882
* 14981: gap of unknown length
* 14982
* 16572: contig of 1591 bp in length
* 16573
* 16672: gap of unknown length
* 18014: contig of 1342 bp in length
* 18015
* 18114: gap of unknown length
* 18115
* 20339: contig of 2225 bp in length
* 20340
* 20439: gap of unknown length
* 20440
* 22021: contig of 1582 bp in length
* 22022
* 22121: gap of unknown length
* 22122
* 24163: contig of 2042 bp in length
* 24164
* 24263: gap of unknown length
* 24264
* 26505: contig of 2242 bp in length
* 26506
* 26605: gap of unknown length
* 26606
* 28894: contig of 2289 bp in length ;
* 28895
* 31231: contig of 2237 bp in length
* 31232
* 31331: gap of unknown length
* 31332
* 32251: contig of 1920 bp in length
* 32252
* 33351: gap of unknown length
* 33352
* 34661: contig of 1310 bp in length
* 34662
* 34761: gap of unknown length
* 34762
* 36350: contig of 1589 bp in length
* 36351
* 36450: gap of unknown length
* 36451
* 38161: contig of 1711 bp in length
* 38162
* 38261: gap of unknown length
* 38262
* 41430: contig of 3169 bp in length
* 41431
* 41530: gap of unknown length
* 41531
* 43197: contig of 1667 bp in length
* 43198
* 43297: gap of unknown length
* 43298
* 45621: contig of 2324 bp in length
* 45622
* 45721: gap of unknown length
* 45722
* 48922: contig of 3201 bp in length
* 48923
* 49023
* 51372: contig of 2350 bp in length
* 51373
* 51472: gap of unknown length
* 51473
* 54730: contig of 3258 bp in length
* 54731
* 54830: gap of unknown length
* 54831
* 57633: contig of 2803 bp in length
* 57634
* 57733: gap of unknown length
* 57734
* 60622: contig of 2889 bp in length
* 60623
* 60722: gap of unknown length
* 60723
* 63956: contig of 3234 bp in length

intron complement(4760..4848)
/gene="F18022_10"
/number=11
exon complement(4849..4922)
/gene="F18022_10"
/number=12
intron complement(4923..5123)
/gene="F18022_10"
/number=12
exon complement(5124..5180)
/gene="F18022_10"
/number=13
intron complement(5181..5288)
/gene="F18022_10"
/number=13
exon complement(5289..5339)
/gene="F18022_10"
/number=14
intron complement(5340..5465)
/gene="F18022_10"
/number=14
exon complement(5466..5531)
/gene="F18022_10"
/number=15
intron complement(5532..5639)
/gene="F18022_10"
/number=15
exon complement(5640..5775)
/gene="F18022_10"
/number=16
intron complement(5776..5929)
/gene="F18022_10"
/number=16
exon complement(5930..5981)
/gene="F18022_10"
/number=17
intron complement(5982..6090)
/gene="F18022_10"
/number=17
exon complement(6091..6118)
/gene="F18022_10"
/number=18
gene 12557..14525
/gene="F18022_20"
CDS join(12557..12652,12752..12889,12984..14525)
/gene="F18022_20"
/note="similarity to 270 kDa ankyrin G isoform, Rattus norvegicus, EMBL:AF102552"
/codon_start=1
/product="ankyrin-like protein"
/protein_id="CAB87762.1"
/db_xref="GI:7573448"
/translation="MGFDLIRPQVAVHALATACGRFVDVVGTVLLKCGVDANSTDRLL
LQSKPSLYTNDVCTALVAIVNRQVSAVRVLLQAGVKTDIMVRLGAMSWDNTGEEF
RVAGVAEPYPLTWCAYEFTSGDILRLLLKQSPNALHNGRTLHHAVLCGKAAV
SVLDDCCADPEAPTKTSGTELPIHIAARDGSVEIIOQLVGFCDINSKNDVGNLAL
LISLTKHPECVKVLADGDAGFGLVKNFGHSVSAIESNKSGLGLERVILELIRFGVV
PHSSNASFVPLLYAGAGDAEALKALYKAQDIYLDQDEGFSAAMLAANGHVEAF
RLVYAGADKLYINSGDTVVVLSSEQNGNRDVIIEKVMLEFALEKDSRNMAGGFYALHC
AARGDVKVALLSGKGYSLDIPDGDTGYTFLMLAAREGHGMCEYLLISCGANCAKNG
RGEKLLDIATDAEKVIRNELSRFRVIEGSSVMKHTKGGKKGKHLRMLLESSGVLS
WGSKRRTVVCKEVIEIGMSQRFKRNKRGKGDABEEGIFRVVTTTNGKVFVFCBGGVLV
CAEMVVRGILVTRETICGRQTQTN"
12557..12652
/gene="F18022_20"
/number=1
intron 12653..12751
/gene="F18022_20"
/number=1
exon 12752..12889
/gene="F18022_20"
/number=2
intron 12890..12983

/gene="F18022_20"
/number=2
12984..14525
/gene="F18022_20"
exon
Alignment Scores:
Pred. No.: 173 Length: 100269
Score: 104.50 Matches: 53
Percent Similarity: 37.55% Conservative: 33
Best Local Similarity: 23.14% Mismatches: 71
Query Match: 10.36% Indels: 72
DB: 8 Gaps: 12
US-09-868-352-23 (1-192) x ATF18022 (1-100269)
Qy 5 ValThrProGluMetLeuAsnTyrAsnLysTyrProGlyProGlnPheIleHisPheGlu 24
Db 58768 ATCACTCCCAACGCTAAATTAT-----TACTCTAAACCCCTTTTATCTGCAAAAGT 58715
Qy 25 AsnIleValLysSerAspIleGluPheGlnLeuValIleAsnGluLysSerAlaPhe 44
Db 58714 AACATTATAAAAAAATGTAACCTCATTTGACAT-----ATAAAGCTCTAACTCGTCTCG 58651
Qy 45 AspValThrValPheGlyGlnArgPheSerGluIleLeuLysTyrAspPheIleVal 64
Db 58660 AATCTCTATAGTTT-----AATATCTTAATCTCTTTTAATTC----- 58622
Qy 65 GlyAspTrpGlyAsnGluGlnLeuArgGlyPheTyrLysAspAlaSer----- 82
Db 58621 -----AGATTGCATAAAAAAACATTTTTCAGGGAAGTAATCTTCAA 58580
Qy 83 ---ThrIleArgLys-----AsnSerArgIleSerArgLeuGluAspTyr 96
Db 58579 TTTTCTGTAGTTAAATAATGGAATTTTAAATAATAAAGCCAAAGCTCGAACTTTA 58520
Qy 97 IleLysGluTyrCysAsnPheGly-----CysAlaTyrPheValLeu 110
Db 58519 ATAAACAGTTTTTGTTCATCTCTTGCTTACCATTTCAAAGTGCACCTTTTCTTCTCA 58460
Qy 111 Glu-----AsnProAsnProArgAspIleLysPhe----- 120
Db 58459 AATCCTCTCTCTCTCTCCGTTCCGAGACAAGCAAAAGCTTCAATCTCAAACTTTCA 58400
Qy 121 -----AspAspGluArgProHisLys 127
Db 58399 GCAGCTACAAAAGTTCTTCTCATCTAAAGACAGTGTGAAGCAGAGAGAGAGAGAG 58340
Qy 128 ArgArgLysSerArg-----SerLysSerGlnSerSerLys 139
Db 58339 AGAGAGAGAGAGATGGAGGTGTGACGTATCATGTTGCGCGAAGTTCGCTTTCTTCC 58280
Qy 140 SerGlnThrArgAsnAsnArgSerGlnSerAsnAlaAsn----- 152
Db 58279 CGCGAGTCCGCTGCTTTACAAAGCTGTTTACGATGAACCTGACCGGACTTCTCTGATGA 58220
Qy 153 AlaHisPheThrSerLysLysArgLysAspThrLysArgGlnGluArgHisIleLys 172
Db 58219 ACCATTTCCTCACCCTG-----AAAAGTTGAAATCCTTAAGCTCCGACACAGTAGAG 58166
Qy 173 GluGluGlnAspLysGluMetThrSer 181
Db 58165 GAACAGAGATCGTGCTATGTACGTCA 58139
RESULT 39
AC120073/c 132934 bp DNA linear HTG 18-JUL-2002
LOCUS Rattus norvegicus clone CH230-357A23, *** SEQUENCING IN PROGRESS
DEFINITION *** 40 unordered pieces.
ACCESSION AC120073
VERSION AC120073.3 GI:21747082
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus

44057T : 44057T

Qy	33	GluPheGlnLeuValIleAsnGluLysSer-----AlaPheAsp-----ValThr	47
Db	32976	GAAATCAGCTGCTCCAGCCCTAAAAAGCTCTTTGCAGAAATTTGACAGACCGCAACATCACT	33035
Qy	48	ValPheGlyGlnArgPheSerGluIleLeuLysTyrAspPheIleValGlyAspTrp	67
Db	33036	GTTAAACACGCAAAACCTTAATGATGCAAAATTTCAAATTTATAGATCATCAAAATTTCTGAATA	33095
Qy	68	GlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAspAlaSerThrIleArgLysAsn	87
Db	33096	TCACCCAGACAAACTGACTATCAGACGATGTCACCGCAT-----	33134
Qy	88	SerArgIleSerArgLeuGluAspTyrIleLysGluTyrCysAsnPheGlyCysAlaTyr	107
Db	33135	CAAAAGCTGGAGATCATAAATAGTATATATTTTCAGACAGAGATCACTACTGGAATAATAATAC	33194

Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talames, J., Tesfaye, S., Theodores, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (27-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 27, 2002 this sequence version replaced gi:18482207.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L21052
Center clone name: 199_E-23
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 201345 bases at least Q40
Consensus quality: 209044 bases at least Q30
Consensus quality: 212024 bases at least Q20
Insert size: 182000; agarose-fp
Insert size: 213377; sum-of-contigs
Quality coverage: 4.6 in Q20 bases; agarose-fp
Quality coverage: 3.9 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 54 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

1 12: contig of 12 bp in length
13 112: gap of 100 bp
113 767: contig of 655 bp in length
768 867: gap of 100 bp
868 1486: contig of 619 bp in length
1487 1586: gap of 100 bp
1587 2581: contig of 995 bp in length
2582 2681: gap of 100 bp
2682 3615: contig of 934 bp in length
3616 3715: gap of 100 bp
3716 4456: contig of 741 bp in length
4457 4556: gap of 100 bp
4557 5548: contig of 992 bp in length
5549 5648: gap of 100 bp
5649 6452: contig of 804 bp in length
6453 6552: gap of 100 bp
6553 7198: contig of 646 bp in length
7199 7298: gap of 100 bp
7299 8131: contig of 833 bp in length
8132 8231: gap of 100 bp
8232 9099: contig of 868 bp in length

9100 9199: gap of 100 bp
9200 10023: contig of 824 bp in length
10024 10123: gap of 100 bp
10124 11265: contig of 1142 bp in length
11266 11365: gap of 100 bp
11366 11997: contig of 632 bp in length
11998 12097: gap of 100 bp
12098 13106: contig of 909 bp in length
13107 14298: contig of 1192 bp in length
14299 14398: gap of 100 bp
14399 15431: contig of 1033 bp in length
15432 15531: gap of 100 bp
15532 16342: contig of 811 bp in length
16343 16442: gap of 100 bp
16443 17510: contig of 1068 bp in length
17511 17610: gap of 100 bp
17611 18866: contig of 1256 bp in length
18867 18966: gap of 100 bp
18967 19993: contig of 1027 bp in length
19994 20093: gap of 100 bp
20094 20938: contig of 845 bp in length
20939 21038: gap of 100 bp
21039 22307: contig of 1269 bp in length
22308 22407: gap of 100 bp
22408 23319: contig of 912 bp in length
23320 23419: gap of 100 bp
23420 24672: contig of 1253 bp in length
24673 24772: gap of 100 bp
24773 25600: contig of 828 bp in length
25601 25700: gap of 100 bp
25701 27029: contig of 1329 bp in length
27030 27129: gap of 100 bp
27130 28673: contig of 1544 bp in length
28674 28773: gap of 100 bp
28774 29795: contig of 1022 bp in length
29796 29895: gap of 100 bp
29896 31567: contig of 1672 bp in length
31568 33129: contig of 1462 bp in length
33130 33229: gap of 100 bp
33230 34638: contig of 1409 bp in length
34639 34738: gap of 100 bp
34739 36991: contig of 2253 bp in length
36992 37091: gap of 100 bp
37092 39530: contig of 2439 bp in length
39531 39630: gap of 100 bp
39631 41804: contig of 2174 bp in length
41805 41904: gap of 100 bp
41905 44637: contig of 2733 bp in length
44638 44737: gap of 100 bp
44738 47882: contig of 3145 bp in length
47883 47982: gap of 100 bp
47983 51826: contig of 3844 bp in length
51827 51926: gap of 100 bp
51927 56275: contig of 4349 bp in length
56276 56375: gap of 100 bp
56376 61809: contig of 5434 bp in length
61810 61909: gap of 100 bp
61910 68212: contig of 6303 bp in length
68213 68312: gap of 100 bp
68313 74425: contig of 6113 bp in length
74426 74525: gap of 100 bp
74526 79707: contig of 5182 bp in length
79708 79807: gap of 100 bp
79808 86515: contig of 6708 bp in length
86516 86615: gap of 100 bp
86616 93246: contig of 6631 bp in length
93247 93346: gap of 100 bp
93347 101225: contig of 7879 bp in length
101226 101325: gap of 100 bp
101326 112385: contig of 11060 bp in length
112386 112485: gap of 100 bp

TITLE
JOURNAL

COMMENT

Db 6019 AAAGAGCAAAAGCGAACGAAGAAAAAACAACAAAGCGCAAGAAAGAAAGAA 6078
QY 136 GlnSerLysSerGlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPhe 155
Db 6079 AAAGAGGAAACAGGAGGAAAAAAGAACACACACACACACACACACACACAC 6138
QY 156 ThrSerLysLysArgLysAspThrLysArgGlnGlnArgHisLysLysGluGln 175
Db 6139 ACAGCGAAAAAGAAAGAAAGGAAAGCAAGAAAGAAAGAAAGAAAGAAAGAA 6198
QY 176 AspLysGluMetThrSerAlaLysGln 184
Db 6199 ACCAAAAAAGAAACAAACAGAA 6225
RESULT 33
LOCUS BC026944 3130 bp mRNA linear ROD 16-AUG-2002
DEFINITION Mus musculus, Similar to splicing factor, arginine/serine-rich 4 (SRP75), clone MGC:30919 IMAGE:4016087, mRNA, complete cds.
ACCESSION BC026944
VERSION BC026944.1 GI:22268150
KEYWORDS MGC.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (04-APR-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: ang@bcm.tmc.edu
Gunsaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M., Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M., Richards, S., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 41 Row: c Column: 13.
Location/Qualifiers
1. 3130
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="C2ECH II"
/clone="MGC:30919 IMAGE:4016087"
/tissue_type="Mammary tumor metastasized to lung."
MWTN-LTR/Wnt1 model. Expression driven by an MWTN-LTR enhancer."
/clone_lib="NCI_CGAP_Lu30"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
992. 2461
/codon_start=1
/product="Similar to splicing factor, arginine/serine-rich 4 (SRP75)"
/protein_id="AAH26944.1"
/db_xref="GI:22268151"
/translation="MAHHIGKAKLTSCWCLAEEDSHVHIQKACGTVLGVRGVEFDD LRADDAVYELNCKDLCSERVIVEHARGPRGSGYGRSGRDKYGPPTPT EYKLIVENLSSRCSDWDLKDYMQAGEVTYDAHGRKKNEGVTFVSYSDMKRALEKL

BASE COUNT 847 a 782 c 862 g 639 t
ORIGIN
Alignment Scores:
Pred. No.: 1.08 Length: 3130
Score: 107.50 Matches: 56
Percent Similarity: 35.10% Conservative: 30
Best Local Similarity: 22.86% Mismatches: 70
Query Match: 10.65% Indels: 89
DB: 10 Gaps: 11
US-09-868-352-23 (1-192) x BC026944 (1-3130)
QY 20 PheIleHisPheGluAsnIleValLysSerAspIleGluPheGlnLeuValIleAsn 39
Db 1106 TTTGTGGAGTTTGATGATCGCTGATGCAGACGATGCTTTTGAAGTGAACGGCAA 1165
QY 40 GluLysSerAlaPheAspValThrVal----- 48
Db 1166 GACCTGTGTGGGAGCGAGTCAATTGTGACACGCTCGAGCGCGGGGATGGCAGC 1225
QY 49 -----PheGlyGlnArgPheSer----- 54
Db 1226 TACGGCTCGGAGCGAGTGGATGATGTTTCGAAGAAGTGGCCGAGATAAATACGGTCC 1285
QY 55 GluIleLeuLeuLysTyrAspPheIleValGlyAsp-----TrpGly 68
Db 1286 CCTACTCGCACACACAGACTTATTGTGGAGAATTGTCAGATCGATCGACGTGGCAG 1345
QY 69 Asn-----GluGlnLeuArgLeuArgGly-----PheTyrLysAspAlaSerThrIle 84
Db 1346 GACCTAAAGGATTACATGCGTCAGGAGAGAGTACGCTATCGATGCTCACAAGGGA 1405
QY 85 ArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyrCysAsnPheGly 104
Db 1406 CGCAAAAC-----GAAGGAGTGATTGAATTTGTCTTACTCT 1444
QY 105 CysAlaTyrPheValLeuGlu-----AsnProAsnProArgAspIleLys 119
Db 1445 GATATGAAAAAGAGCTTTGGAAAAAGCTGGAGGAACTGAAGTCAACGACAGAAATCAGA 1504
QY 120 PheAspAspLysArgPro----- 126
Db 1505 TTAGTTGAGACAAAGCCAGGTTCTAGCGCGCGGCTTTTACTCCAGAACCGGAGTCCAC 1564
QY 127 LysArg-----ArgLysSerArgSerLysSerGlnSerSer 138
Db 1565 TCAAGGTCTCGTCTCGAAGCAGACATTCCTCCGGAAGACAGACGCGGAGTGGTAGCAGT 1624
QY 139 LysSerGlnThrArgAsnAsnArgSerGlnSerAsnAlaHisPheThrSerLys 158
Db 1625 AAAAGACGACATTCGAAGAGCCGCTCGATCCAGGTTCAGGTTCCTCCACTCCCGCAGCAAG 1684
QY 159 LysArg----- 160
Db 1685 AGCCGAGTGGCAGTCAGACCGCCAGCTAAGAGGAGAAAAAGCCGAGCCCCCAGC 1744
QY 161 LysAspThrLysArgArgGlnGluArgHisIleLysGluGlnAspLysGluMetThr 180
Db 1745 AAGCAACAAGACCGCCAGTCCG-----AGCCGAGCCCTGACACAGACCGCAGT 1795
QY 181 SerAlaLysGlnHis 185
Db 1796 AAGAGTAAAGACCAAC 1810
RESULT 34
ALS90994

RESULT 32

```

AC083856          215455 bp      DNA      linear      HTG 04-OCT-2000
LOCUS             Mus musculus chromosome 12 clone RP23-366H13 strain C57BL6/J,
DEFINITION        WORKING DRAFT SEQUENCE, 16 unordered pieces.
ACCESSION         AC083856
VERSION           AC083856.1 GI:10567854
KEYWORDS          HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE            Mus musculus
ORGANISM          Mus musculus
REFERENCE          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 215455)
Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W.,
Bouffard, G.G., Dietrich, N.L., Gupta, J., Ho, S.-L., Idol, J.,
Lee-Lin, S.-G., Legaspi, R., Lim, M., Maduro, Q.L., Maduro, V.B.,
Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Snyder, B.,
Stantripop, S., Thomas, J.W., Thomas, P.J., Tingson, E.E.,
Touchman, J.W., Tran, J.T., Vogt, J.L., Walker, M.A., Wetherby, K.D. and
Green, E.D.
TITLE             NISC Mouse Sequencing Initiative
JOURNAL           Unpublished
REFERENCE         2 (bases 1 to 215455)
AUTHORS           Green, E.D.
TITLE             Direct Submission
JOURNAL           Submitted (04-OCT-2000) NIH Intramural Sequencing Center, 8717
COMMENT           Grovemont Circle, Gaithersburg, MD 20877, USA
----- Genome Center -----
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.mouse@phgrl.nih.gov
----- Project Information -----
Center project name: sh
Center clone name: 366H13
----- Summary Statistics -----
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 207240 bases at least Q40
Consensus quality: 209922 bases at least Q30
Consensus quality: 211032 bases at least Q20
Insert size: 214000; agarose-fp
Insert size: 229000; pulse-field-gel
Insert size: 213955; sum-of-contigs
Quality coverage: 5.63x in Q20 bases; agarose-fp
Quality coverage: 5.26x in Q20 bases; pulse-field-gel
Quality coverage: 5.63x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 5773: contig of 5772 bp in length
* 5873: gap of unknown length
* 10133: contig of 4261 bp in length
* 10134: 10233: gap of unknown length
* 10234: 15586: contig of 5353 bp in length
* 15587: 15686: gap of unknown length
* 15687: 20246: contig of 4560 bp in length
* 20247: 20346: gap of unknown length
* 20347: 26852: contig of 6506 bp in length
* 26853: 26952: gap of unknown length
* 26953: 36105: contig of 9153 bp in length
* 36106: 36205: gap of unknown length
* 36206: 44497: contig of 8292 bp in length
* 44498: 44597: gap of unknown length
* 44598: 53137: contig of 8540 bp in length
* 53138: gap of unknown length

```

```

* 53238 62667: contig of 9430 bp in length
* 62668 62767: gap of unknown length
* 62768 71276: contig of 8509 bp in length
* 71277 71376: gap of unknown length
* 71377 81084: contig of 9708 bp in length
* 81085 81184: gap of unknown length
* 81185 95747: contig of 14563 bp in length
* 95748 95847: gap of unknown length
* 95848 112911: contig of 17064 bp in length
* 112912 113011: gap of unknown length
* 113012 132279: contig of 19268 bp in length
* 132280 132379: gap of unknown length
* 132380 159697: contig of 27318 bp in length
* 159698 159797: gap of unknown length
* 159798 215455: contig of 55658 bp in length.
FEATURES          Location/Qualifiers
source            1..215455
                  /organism="Mus musculus"
                  /strain="C57BL6/J"
                  /db_xref="taxon:10090"
                  /chromosome="12"
                  /clone_lib="RPCI mouse BAC library 23"
                  /clone="RP23-366H13"
misc_feature      1..5772
                  /note="assembly_fragment"
misc_feature      5873..10133
                  /note="assembly_fragment"
                  /note="end:SP6"
                  /vector_side:right"
misc_feature      10234..15586
                  /note="assembly_fragment"
misc_feature      15687..20246
                  /note="assembly_fragment"
                  /note="end:T7"
                  /vector_side:right"
misc_feature      20347..26852
                  /note="assembly_fragment"
misc_feature      26953..36105
                  /note="assembly_fragment"
misc_feature      36206..44497
                  /note="assembly_fragment"
misc_feature      44598..53137
                  /note="assembly_fragment"
misc_feature      53238..62667
                  /note="assembly_fragment"
misc_feature      62768..71276
                  /note="assembly_fragment"
misc_feature      71377..81084
                  /note="assembly_fragment"
misc_feature      81185..95747
                  /note="assembly_fragment"
misc_feature      95848..112911
                  /note="assembly_fragment"
misc_feature      113012..132279
                  /note="assembly_fragment"
misc_feature      132380..159697
                  /note="assembly_fragment"
misc_feature      159798..215455
                  /note="assembly_fragment"
BASE COUNT        68411 a 40237 c 39791 g 65506 t 1510 others
ORIGIN
Alignment Scores:
Pred. No.:        213
Score:            108.00
Percent Similarity: 60.87%
Best Local Similarity: 28.99%
Query Match:      10.70%
DB:               2
Matches:          215455
Conservative:     20
Mismatch:         27
Indels:           0
Gaps:             0
US-09-868-352-23 (1-192) x AC083856 (1-215455)
QY 116 ArgAspIleLysPheAspAspGluArgProHisLysArgArgLysSerArgSerLysSer 135

```


BC019437.1 GI:18043895
MGC.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 2204)
Strausberg, R.
Direct Submission
Submitted (19-DEC-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Baylor College of Medicine Human Genome
 Sequencing Center

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: ang@bcm.tmc.edu
Gunarane, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAK Plate: 41 Row: d Column: 19
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 10181125.
 Location/Qualifiers
 1..2204

```

/db_xref="taxon:10090"

```

```
/clone="MGC:30385 IMAGE:4018178"
```

arose spontaneously from a senescent normal mammary (clonal) outgrowth infected with the virus MMTV." /clone_lib="NCI CGAP Lu29"

```
/note="Vector: pCMV-SPORT6"
```

```
cc: i386
/codon start=1
```

```

/protein_id="AAH19437.1"
/db_xref="GI:18043896"
/translation="MPRVYIGRLSYOAREEDVERFFKGVCKILVDFIKNGYGEVEFD
```

ERDADDAV I EELNORDECGERV I VEHARGF AKNDGSI GSGKSGTGI KKSGRDKI GFF YIKI
 EYBLI TVENI.SSRCSWODLKDYMBRAGEVTYADAHKGRKNEGVTIEFVCSYSDMKRAL.EKI.

DGEVNGKKIKLVEDRFGSSKKKKKSIKSKSHSKSKSKSHSKSKSGSSKSSHKS
 BSRSPSGSHSRSKSRSPSOSRSPSRSKKFKSRSPSKDNKSRSPSRSPDKSRSKSDHAFD

KLQ NND SAGNARKSHF SKIDSKRSRQ ERKAELEKKKV ANAKSQERKSQEKSEBK
SRSSRSSRSRKDKRGKRSPDESRSRSRSRSRSRSRSRSRSRSRSRSRSRSRSRSRS

ANDIDH3K5F3K5V3R6K6H6A6E6Q6R6S6A6E6S6A6F6N6F6N6K6N6S6I6S6K6N6V6
AF6R6S6R6S6K6A6S6K6T6R6S6R6K6S6P6R6S6A6S6P6S6R6S6R6S6H6S6R6S"

2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030 2031 2032 2033 2034 2035 2036 2037 2038 2039 2040 2041 2042 2043 2044 2045 2046 2047 2048 2049 2050 2051 2052 2053 2054 2055 2056 2057 2058 2059 2060 2061 2062 2063 2064 2065 2066 2067 2068 2069 2070 2071 2072 2073 2074 2075 2076 2077 2078 2079 2080 2081 2082 2083 2084 2085 2086 2087 2088 2089 2090 2091 2092 2093 2094 2095 2096 2097 2098 2099 2100 2101 2102 2103 2104 2105 2106 2107 2108 2109 2110 2111 2112 2113 2114 2115 2116 2117 2118 2119 2120 2121 2122 2123 2124 2125 2126 2127 2128 2129 2130 2131 2132 2133 2134 2135 2136 2137 2138 2139 2140 2141 2142 2143 2144 2145 2146 2147 2148 2149 2150 2151 2152 2153 2154 2155 2156 2157 2158 2159 2160 2161 2162 2163 2164 2165 2166 2167 2168 2169 2170 2171 2172 2173 2174 2175 2176 2177 2178 2179 2180 2181 2182 2183 2184 2185 2186 2187 2188 2189 2190 2191 2192 2193 2194 2195 2196 2197 2198 2199 2200 2201 2202 2203 2204 2205 2206 2207 2208 2209 2210 2211 2212 2213 2214 2215 2216 2217 2218 2219 2220 2221 2222 2223 2224 2225 2226 2227 2228 2229 2230 2231 2232 2233 2234 2235 2236 2237 2238 2239 2240 2241 2242 2243 2244 2245 2246 2247 2248 2249 2250 2251 2252 2253 2254 2255 2256 2257 2258 2259 2260 2261 2262 2263 2264 2265 2266 2267 2268 2269 2270 2271 2272 2273 2274 2275 2276 2277 2278 2279 2280 2281 2282 2283 2284 2285 2286 2287 2288 2289 2290 2291 2292 2293 2294 2295 2296 2297 2298 2299 2300 2301 2302 2303 2304 2305 2306 2307 2308 2309 2310 2311 2312 2313 2314 2315 2316 2317 2318 2319 2320 2321 2322 2323 2324 2325 2326 2327 2328 2329 2330 2331 2332 2333 2334 2335 2336 2337 2338 2339 2340 2341 2342 2343 2344 2345 2346 2347 2348 2349 2350 2351 2352 2353 2354 2355 2356 2357 2358 2359 2360 2361 2362 2363 2364 2365 2366 2367 2368 2369 2370 2371 2372 2373 2374 2375 2376 2377 2378 2379 2380 2381 2382 2383 2384 2385 2386 2387 2388 2389 2390 2391 2392 2393 2394 2395 2396 2397 2398 2399 2400 2401 2402 2403 2404 2405 2406 2407 2408 2409 2410 2411 2412 2413 2414 2415 2416 2417 2418 2419 2420 2421 2422 2423 2424 2425 2426 2427 2428 2429 2430 2431 2432 2433 2434 2435 2436 2437 2438 2439 2440 2441 2442 2443 2444 2445 2446 2447 2448 2449 2450 2451 2452 2453 2454 2455 2456 2457 2458 2459 2460 2461 2462 2463 2464 2465 2466 2467 2468 2469 2470 2471 2472 2473 2474 2475 2476 2477 2478 2479 2480 2481 2482 2483 2484 2485 2486 2487 2488 2489 2490 2491 2492 2493 2494 2495 2496 2497 2498 2499 2500 2501 2502 2503 2504 2505 2506 2507 2508 2509 2510 2511 2512 2513 2514 2515 2516 2517 2518 2519 2520 2521 2522 2523 2524 2525 2526 2527 2528 2529 2530 2531 2532 2533 2534 2535 2536 2537 2538 2539 2540 2541 2542 2543 2544 2545 2546 2547 2548 2549 2550 2551 2552 2553 2554 2555 2556 2557 2558 2559 2560 2561 2562 2563 2564 2565 2566 2567 2568 2569 2570 2571 2572 2573 2574 2575 2576 2577 2578 2579 2580 2581 2582 2583 2584 2585 2586 2587 2588 2589 2590 2591 2592 2593 2594 2595 2596 2597 2598 2599 2600 2601 2602 2603 2604 2605 2606 2607 2608 2609 2610 2611 2612 2613 2614 2615 2616 2617 2618 2619 2620 2621 2622 2623 2624 2625 2626 2627 2628 2629 2630 2631 2632 2633 2634 2635 2636 2637 2638 2639 2640 2641 2642 2643 2644 2645 2646 2647 2648 2649 2650 2651 2652 2653 2654 2655 2656 2657 2658 2659 2660 2661 2662 2663 2664 2665 2666 2667 2668 2669 2670 2671 2672 2673 2674 2675 2676 2677 2678 2679 2680 2681 2682 2683 2684 2685 2686 2687 2688 2689 2690 2691 2692 2693 2694 2695 2696 2697 2698 2699 2700 2701 2702 2703 2704 2705 2706 2707 2708 2709 2710 2711 2712 2713 2714 2715 2716 2717 2718 2719 2720 2721 2722 2723 2724 2725 2726 2727 2728 2729 2730 2731 2732 2733 2734 2735 2736 2737 2738 2739 2740 2741 2742 2743 2744 2745 2746 2747 2748 2749 2750 2751 2752 2753 2754 2755 2756 2757 2758 2759 2760 2761 2762 2763 2764 2765 2766 2767 2768 2769 2770 2771 2772 2773 2774 2775 2776 2777 2778 2779 2780 2781 2782 2783 2784 2785 2786 2787 2788 2789 2790 2791 2792 2793 2794 2795 2796 2797 2798 2799 2800 2801 2802 2803 2804 2805 2806 2807 2808 2809 2810 2811 2812 2813 2814 2815 2816 2817 2818

cores:

0:333	Length:	2204
108.50	Matches:	57

Similarity: 22.98%
 Conservativity: 30
 Mismatches: 72

10:758	Index:	63
10	Gaps:	11

52-23 (1-192) x BC019437 (1-2204)

[vprG]nphET[eh]isphoG[Asn]T[eva]l[vsce]rAsnAsnT[eG]nphoG[nTe] 36

Qy 18 ---ProGlnPheIleHisPheGluAsnIleValLysSerAspAspIleGluPheGlnLeu 36
|||||...::: ||| ::::: |||
Db 1220 TTACGCGAGTACCAATACCAAGAGCTTATGTTATCAAGAAAGGTCCTTAACGCA 1279

Qy 37 VallleAsnGluLysSerAlaPheAspValThrValPheGlyGlnArgPheSerGluIle 56
|||||...::: ||| ::::: |||
Db 1280 ACGATAAATCCAGAGAGC-----AGGGAAT 1306

Qy 57 LeuLeu-----LysTyr-----Asp 61
::: |||
Db 1307 GTGATGCCCGCTACTATAATATACGTGATAGAGAAGCAGCTTCAGTAATAATCC 1366

Qy 62 PheIleValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAspAla 81
|||||...::: ||| ::::: |||
Db 1367 TTCATTGTT---GATCAAGGACAA-----ATTACAGGCGCT 1402

Qy 82 SerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGlu----- 99
|||||...::: ||| ::::: |||
Db 1403 TCCACTTTTCACTAAGCAGCTT-AATCAAAATGTTTTTCAAGATTACTTTACAGAGGCGATC 1461

Qy 100 -----TyrCysAsnPheGlyCys-----AlaTyrPhe-Va 109
|||
Db 1462 TTCGAAATGGGTGCACCTTGGTGTGTCGTAACCGGAATATTACGGATGTATGGTTAAAC 1521

Qy 109 lLeuGluAsnProAsnProArgAspIleLysPheAspAspGluArgProHisLysArgAr 129
::: |||...::: ||| ::::: |||
Db 1522 GATGGAATGGCGCAAGATATG---TTAAATATATAAGAAAGATGAACCTGCTGAGCAAG 1578

Qy 129 gLysSerArgSerLysSerGlnSer-----SerLy 139
::: |||...::: ||| ::::: |||
Db 1579 ACGTTCCAACAGTGCATGCTTCATCCCGCAGAAAGTGAAAGAAATAACGTCAAGGTAA 1638

Qy 139 sSerGlnThrArgAsnArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysLy 159
|||||...::: ||| ::::: |||
Db 1639 ATCATCGGTAAGAGTAACTCAACCTCAACGATGCTTCTTAACAAAAGAAACCAAGAAAA 1698

Qy 159 sArgLysAspThrLysArgGlnGlnArgHisIleLysGluGluGlnAspLysGlu 178
|||||...::: ||| ::::: |||
Db 1699 AAGAAAGCCACCGAAGTAATGATCTGAGATAATCTTCGGAAGAAAGACAAGAG 1756

RESULT 29
SCXII
LOCUS
DEFINITION S.cerevisiae 14.6kb fragment from chromosome VII.
ACCESSION X92670
VERSION X92670.1 GI:1246838
KEYWORDS ribosomal protein.
SOURCE Saccharomyces cerevisiae.
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE
AUTHORS Escribano,V., Eraso,P., Portillo,F. and Mazon,M.J.
TITLE Sequence analysis of a 14.6 kb DNA fragment of Saccharomyces
cerevisiae chromosome VII reveals SEC27, SSM1b, a putative
S-adenosylmethionine-dependent enzyme and six new open reading
frames
JOURNAL Yeast 12 (9), 887-892 (1996)
MEDLINE 96437978
PUBMED 8840506
REFERENCE
AUTHORS Mazon,M.J.
TITLE Direct Submission
JOURNAL Submitted (31-OCT-1995) M.J. Mazon, C.S.I.C., Instituto
Investigaciones Biomedicas, Arturo Duperier 4, E-28029 Madrid,
Spain
FEATURES
source Location/Qualifiers
1. 14607
/organism="Saccharomyces cerevisiae"
/strain="FY1679"
/db_xref="taxon:4932"
/chromosome="VII"
1. 566
gene

/gene="G2812"
<1. 566
/gene="G2812"
/codon_start=3
/protein_id="CAA63357.1"
/db_xref="GI:1246839"
/db_xref="SWISS-PROT:P53121"
/translation="SVANELLAALGNVAKHDDNSDYESNDTGVNDELKQAODETPTPT
VTSDDNKPTFSEKILSKFSPKNEASTDALVERPKQOTFPHNLNLSRENLSLIG
KDPYPGHTRSQSDAHNKLINSFEEEDTSNTDFHDSTEGDLLDTSSSGGFRSQNTV
RDSISNLGNKQPLRRKPPGFDEGFM"
complement(936. 1973)
/gene="G2816"
complement(936. 1973)
/gene="G2816"
/codon_start=1
/protein_id="CAA63358.1"
/db_xref="GI:1246840"
/db_xref="SWISS-PROT:P53122"
/translation="MYNLFEEFFFIYITLTKSTFVGKSSFPGHDVCYKFDQNFQTE
FFNLVKGDKLQNLKOEYEOYKKOSTLYTGVEIKOYEOVAPLOINNFLOVTFCKGG
KPTNHLIPQKOLDNRAEPLCIPAQEDDTISQSSVCFKARVKYTORNVLYFPNK
FVGFEVCNSKTHPTQNNFETIPLPIISDDIRDKISWSINGRITKVVVYLHSUS
IPMSKYMELIHSKNTSLELYKSLTSEFWKSYKLNKFKKEMKSWINEPNMEQYDS
SGNENVPNFEHKLDRTINRIAQNIERPHDALIRAIHAHRNSNSGNITRRKYLRRKIS
KMLKNKTLPLKN"
join(2311. 2328, 2529. 5180)
/gene="G2827"
/note="SEC27"
join(2311. 2328, 2529. 5180)
/gene="G2827"
/note="see also U11237"
yeast coatomer subunit"
/codon_start=1
/protein_id="CAA63359.1"
/db_xref="GI:1246841"
/db_xref="SWISS-PROT:P41811"
/translation="MKLDIKKTSNRSDRVKGIDFHPTPEWVLTLYSGRVELWNYET
QVRSIQVTETPVKAGFKARKNWILVSGDDPRIRVFNNTGKGVYDFEAHPDYIRS
IAVHTKPYVLSGSDLLTVKLNWENNWALEQIFEGHEFVCMVFANPKDPTFASGC
LDRTVKVLSQSTPNFTLTGGQGVNVDYYPDKPYMTASDDLTIKINDYQTK
SCVATLEGHMSVSEAFVHTPLPIIISGSEDGTLKIWNSTYKVEKTLNGLERSWC1
ATHTGRKQNIASGDFNGFTVLSLGNDEPTLSLDPPVKLVMSGGKNAASDIFTAVIR
GNEVEQDEPLSLQTKELGSDVDFPOSLSHSPNGRFVTVVGDEYVLYTALAWRNKAF
GKCODEPWCSPDSYALIDETGOIKYKNEKEVTSNVPMSAIDRLFSALLGVKSD
GFYFDDNGTLVLRIDVNAKDIWSDNGELVMVNTNSNGDEASGTTLLFNKDATL
EAAANNIDDSQVDEAFDVLIELSESLTSGKWGVDFIFTATNLNVFVGKTYNL
HYTKEMYLIGYLARDNKVYLADREHVHYGEISLEVLEFQTLTLRGEI BEA1ENVLP
NVEGKSLTKIARLEQOYEEALNISPQDKFELAKVGOLTLARLDTDESADM
KWRALGDASLORENFKLAVEAFTNAHDESLEFLHSSFNKKEGLVTLAKDAERAGFN
LAFNAYIAGIDOGARDLLIKSORFSEAAFLSTYGLGDDAVNDIVTKWENLILNK
NIVSERVCGAELPGSSGDAQPLDLDSTPAPEQADENKREAVEDESEFKNSEAV
EAEKKEEAPQQQSEQQPEQGEAVPEPVEES"
2329. 2528
/gene="G2827"
complement(5339. 6301)
/gene="G2830"
complement(5339. 6301)
/gene="G2830"
/codon_start=1
/protein_id="CAA63360.1"
/db_xref="GI:1246842"
/db_xref="SWISS-PROT:P53123"
/translation="MILVYNRIRSISSSLGRTHVRVNSNSQNRWLNRLQKDPYTKKA
KVGNLRSRAAFKLMQIDDKYRFLFSKNRDTQRIIDLGYAPGASQVARSQSPSMILG
VDLLPCPEGHVNSIQANTLAKRTHDLIRLFFSKHFLNHRDHLKHDGTFORMLLEE
LTHVKDTELYREIETSDDIYETPTNTSIREKFPVDVI ESDMYEPWPOTGTGFWNNI
TNOAYFRMANTSGVSTRDHYQSIDLCDAALVTADLLRLPLGSFVCKLYLTGTEENLFPKK
RQAVFTNVHKFPKASRDESKETYIIGLKKRNVDKLDVFSN"
7083. 7736
/gene="G2834"
/note="SSM1b"
7083. 7736
intron
gene
CDS


```

DEFINITION Sequence 7936 from Patent WO0229113.
ACCESSION AX439521
VERSION AX439521.1 GI:216664332
KEYWORDS
SOURCE
ORGANISM
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE
1
AUTHORS Berka R. and Clausen, I.G.
TITLE Methods for monitoring multiple gene expression
JOURNAL Patent: WO 0229113-A 7936 11-APR-2002;
Novozymes Biotech, Inc. (US) ; Novozymes A/S (DK)
FEATURES
Location/Qualifiers
1..285
/organism="Bacillus clausii"
/db_xref="taxon:79880"
BASE COUNT 93 a 51 c 66 g 74 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 6.92e-14 Length: 285
Score: 232.50 Matches: 46
Percent Similarity: 69.32% Conservative: 15
Best Local Similarity: 52.27% Mismatches: 22
Query Match: 23.04% Indels: 5
Gaps: 2
DB: 6

US-09-868-352-23 (1-192) x AX439521 (1-285)
Qy 26 ileVallySerAspAspIleGluPheGlnLeuValIleAsnGluLysSerAlaPheAsp 45
Db 1 ATGGTGACGATTTCAGGCAATTGAATATGATCAAGTCCTCGAAACGAGCGCANTGGTGAAT 60
Qy 46 ValThrValPheGlyClnArgPheSerGlnLeuLeuLysTyrAspPheIleValGly 65
Db 61 GAGGAGGCAATCAACGCGGTACGCGATGACTGTAACAATATGATTACATCGTCGCGC 120
Qy 66 AspTrpGlyAsnGluInLeuArgLeuArgGlyPheTyrLysAsp-----AlaSer 82
Db 121 GATTGGGGCTACACCACTCGGTTGCGCGCTTTTGTGATGACAAAACAAAAATCA 180
Qy 83 ThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyrCysasn 102
Db 181 ACATTCGATAAA-----AAATAGCACGTATCCCGACTATCTTACGAATACATGCAAT 234
Qy 103 PheGlyCysAlaTyrPheValLeu 110
Db 235 TTGGTTCGCGCTATTGTTGCTTG 258

RESULT 24
BSUB0017/c
LOCUS
DEFINITION BSUB0017 217420 bp DNA linear BCT 26-NOV-1997
Bacillus subtilis complete genome (section 17 of 21): from 3197001
to 3414420.
ACCESSION 299120 AL009126
VERSION 299120.1 GI:2635613
KEYWORDS
SOURCE
ORGANISM
Bacillus subtilis.
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE
1 (bases 1 to 217420)
AUTHORS Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G.,
Azevedo, V., Bertero, M.G., Bessieres, P., Bolotin, A., Borchert, S.,
Borries, R., Boursier, L., Brans, A., Braun, M., Brignell, S.C.,
Bron, S., Brouillet, S., Bruschi, C.V., Caldwell, B., Capuano, V.,
Carter, N.M., Choi, S.K., Codani, J.J., Connerton, I.F., Cummings, N.J.,
Daniel, R.A., Denizot, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D.,
Emmerson, P.T., Entian, K.D., Errington, J., Fabret, C., Ferrari, E.,
Faulstich, D., Fritz, M., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A.,
Galleron, N., Ghim, S.Y., Glaser, P., Goffeau, A., Golightly, E.J.,
Grandi, G., Guiseppli, G., Guy, B.J., Haga, K., Haeck, J., Harwood, C.R.,
Henauf, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.F.,
Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y.,
Klaerr-Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P.,
Koningstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A.,
Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H.,
Masuda, S., Maugel, C., Medigue, C., Medina, N., Mellado, R.P.,
Mizuno, M., Moesti, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M.,
Ogawa, K., Ogiwara, A., Oudega, B., Park, S.H., Parro, V., Pohl, T.M.,
Portetelle, D., Porwollik, S., Prescott, A.M., Presecan, E., Pujic, P.,
Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M.,
Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadate, Y., Sato, T.,
Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J.,
Sekowska, A., Seror, S.J., Serror, P., Shin, B.S., Soldo, B.,
Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K.,
Takeuchi, M., Tamakoshi, A., Tanaka, T., Terpstra, P., Tognoni, A.,
Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassarotti, A.,
Viari, A., Wambutt, R., Wedler, E., Wedler, H., Weitznegger, T.,
Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K.,
Yoshida, K., Yoshikawa, H.F., Zumbstein, E., Yoshikawa, H. and
Danchin, A.
The complete genome sequence of the gram-positive bacterium
Bacillus subtilis
Nature 390 (6657), 249-256 (1997)
98044033
9384377
2 (bases 1 to 217420)
Direct Submission
Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.
Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,
Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724
Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,
adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
68 89 48
FEATURES
Location/Qualifiers
1..217420
/organism="Bacillus subtilis"
/strain="168"
/db_xref="taxon:1423"
complement(3..1277)
/gene="yulE"
complement(3..1277)
/gene="yulE"
/function="unknown"
/notes="similar to L-rhamnose isomerase"
/codon_start=1
/transl_table=11
/protein_id="CAB15107.1"
/db_xref="GI:2635614"
/db_xref="SPTREMBL:O05264"
/translation="MTIKANYDSAKQAYEKWGDVVEALRQLEQVPISIHWCQDDITE
GFVNKGLSGGIDVTGNYPGKATPEELRDLKALSLIPGKHVNLHAIYAETNRE
AVERDELKPOHFENWVKWAKNLGLDFNPFLFSBKADGLTSLPDDPDIRSEFIRH
CIACRIGYFGKELGTPCLTNWIIPDGKIDIPSDRLTPKRLKESLDLRFSEISQ
HNLDSESLKFLGSGESYVVGSHFEYLALAYLTNKKLCLLDLTGHFHPPTVSNKISSML
LYTDLKALHVSFVRWSDSHVVDDELREIALEIVRNHAKLEIVAGLGLDFDASINRV
AAWTIGTRNMKALYALLPGLYKQLQEGREYTERLALMBEFTKYPFGAINDSYCE
QMGVPVKEAWLYDIKEYEQVLLKRASSPIV"
complement(1303..1617)
/gene="yulD"
complement(1303..1617)
/gene="yulD"
/function="unknown"
/notes="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB15108.1"
/db_xref="GI:2635615"
/db_xref="SPTREMBL:O05263"
/translation="MKRKASIMFVHQDYEEYKQRHDDIMPENAEALKARGAHYSTIF
LDSETRGLFAYLEIDEEKRWKMDTEVCQRWKNMAPLTKNTSDFSPVAIDLKEYV
LD"
complement(1635..3092)
/gene="yulC"
complement(1635..3092)
/gene="yulC"

```



```

Qy 108 PheValLeuGlu 111          2690 bp  DNA  linear  PAT 14-JUN-2002
|||||:|||||
Db 775 TTTGTTATTAAA 786

RESULT 21
AX416642      AX416642      2690 bp  DNA  linear  PAT 14-JUN-2002
LOCUS      Sequence 3633 from Patent WO0228891.
DEFINITION
ACCESSION  AX416642
VERSION    AX416642.1  GI:21449099
KEYWORDS
SOURCE    Listeria monocytogenes ATCC 19115.
ORGANISM  Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
REFERENCE
AUTHORS   Glaser, P. and Kunst, F.
TITLE     Listeria innocua, genome and applications
JOURNAL   Patent: WO 0228891-A 3633 11-APR-2002;
          Pasteur Institut (FR)
FEATURES
  source   1..2690
            /organism="Listeria monocytogenes ATCC 19115"
            /db_xref="taxon:176281"
BASE COUNT  852 a 456 c 629 g 753 t
ORIGIN

Alignment Scores:
Pred. No.:      2.61e-13      Length:      2690
Score:          239.50      Matches:      43
Percent Similarity: 77.38%      Conservative: 22
Best Local Similarity: 51.19%      Mismatches:  14
Query Match:    23.74%      Indels:       5
DB:             6           Gaps:         2

US-09-868-352-23 (1-192) x AX416642 (1-2690)

Qy 31 AspileGluPheGlnLeuValIleAsnGluLysSerAlaPheAspValThrValPheGly 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 541 GATTAAACTACGAGATTATCACCATTATCCGACGGCTTTTGACGGAGGAAGTTAAAT 600

Qy 51 GlnArgPheSerGluIleLeuLeuLysTyrAspPheIleValGlyAspTrpGlyAsnGlu 70
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 601 GAACGGTTTAGCGATATCTTTGGCGGATATGATTATATAGTAGCGACTGGGGTTACGAT 660

Qy 71 GlnLeuArgLeuArgGlyPheTyrLysAspAlaSerThrIleArgLys-----Asn 87
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 661 CAATCCGCTCTCAAGGCGTTTTTTTGAAGACGATAAC-----CGCAAAGCCGATACGAC 714

Qy 88 SerArgIleSerArgLeuGluAspTyrIleLysGluTyrCysAsnPheGlyCysAlaTyr 107
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 715 AACAAATTACACGTTAAAGATATATTTACGAATACTGCAATTTTGGTTGTCGTAT 774

Qy 108 PheValLeuGlu 111
|||||:|||||
Db 775 TTTGTTATTAAA 786

RESULT 22
AP001518/c    AP001518      303250 bp  DNA  linear  BCT 10-JAN-2001
LOCUS      Bacillus halodurans genomic DNA, section 12/14.
DEFINITION
ACCESSION  AP001518  BA000004
VERSION    AP001518.1  GI:10175792
KEYWORDS
SOURCE    Bacillus halodurans DNA.
ORGANISM  Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE
AUTHORS   Takami, H., Nakasone, K., Hirama, C., Takaki, Y., Masui, N., Fuji, F.,
          Nakamura, Y. and Inoue, A.
TITLE     An improved physical and genetic map of the genome of alkaliphilic
          Bacillus sp. C-125
JOURNAL   Extremophiles 3 (1), 21-28 (1999)
MEDLINE   99184645

```

```

PUBMED
REFERENCE
AUTHORS   Takami, H. and Horikoshi, K.
TITLE     Reidentification of facultatively alkaliphilic Bacillus sp. C-125
          to Bacillus halodurans
JOURNAL   Biosci. Biotechnol. Biochem. 63, 943-945 (1999)
REFERENCE
AUTHORS   Takami, H., Nakasone, K., Ogasawara, N., Hirama, C., Nakamura, Y.,
          Masui, N., Fuji, F., Takaki, Y., Inoue, A. and Horikoshi, K.
TITLE     Sequencing of three lambda clones from the genome of alkaliphilic
          Bacillus sp. strain C-125
JOURNAL   Extremophiles 3 (1), 29-34 (1999)
MEDLINE   99184646
PUBMED   10086842
REFERENCE
AUTHORS   Takami, H., Takaki, Y., Nakasone, K., Hirama, C., Inoue, A. and
          Horikoshi, K.
TITLE     Sequence analysis of a 32-kb region including the major ribosomal
          protein gene clusters from alkaliphilic Bacillus sp. strain C-125
JOURNAL   Biosci. Biotechnol. Biochem. 63 (2), 452-455 (1999)
MEDLINE   99209008
PUBMED   10192928
REFERENCE
AUTHORS   Takami, H., Takaki, Y., Nakasone, K., Sakiyama, T., Maeno, G.,
          Sasaki, R., Hirama, C., Fuji, F. and Masui, N.
TITLE     Genetic analysis of the chromosome of alkaliphilic Bacillus
          halodurans C-125
JOURNAL   Extremophiles 3 (3), 227-233 (1999)
MEDLINE   99411980
PUBMED   10484179
REFERENCE
AUTHORS   Takami, H.
TITLE     Genome analysis of facultatively alkaliphilic Bacillus halodurans
          C-125
JOURNAL   (in) Extremophiles in deep-sea environments (Ed.),
          : 249-284: Springer-Verlag (1999),
          7 (sites)
REFERENCE
AUTHORS   Takami, H., Masui, N., Nakasone, K. and Horikoshi, K.
TITLE     Replication origin region of the chromosome of alkaliphilic
          Bacillus halodurans C-125
JOURNAL   Biosci. Biotechnol. Biochem. 63 (6), 1134-1137 (1999)
MEDLINE   99356711
PUBMED   10427704
REFERENCE
AUTHORS   Takami, H. and Horikoshi, K.
TITLE     Analysis of the genome of an alkaliphilic Bacillus strain from an
          industrial point of view
JOURNAL   Extremophiles 4 (2), 99-108 (2000)
MEDLINE   20263314
PUBMED   10805564
REFERENCE
AUTHORS   Nakasone, K., Masui, N., Takaki, Y., Sasaki, R., Maeno, G., Sakiyama, T.,
          Hirama, C., Fuji, F. and Takami, H.
TITLE     Characterization and comparative study of the rrm operons of
          alkaliphilic Bacillus halodurans C-125
JOURNAL   Extremophiles 4 (4), 209-214 (2000)
MEDLINE   20426005
PUBMED   10972189
REFERENCE
AUTHORS   Takami, H., Nakasone, K., Takaki, Y., Maeno, G., Sasaki, R., Masui, N.,
          Fuji, F., Hirama, C., Nakamura, Y., Ogasawara, N., Kuhara, S. and
          Horikoshi, K.
TITLE     Complete genome sequence of the alkaliphilic bacterium Bacillus
          halodurans and genomic sequence comparison with Bacillus subtilis
          Nucleic Acids Res. 28 (21), 4317-4331 (2000)
MEDLINE   20512582
PUBMED   11058132
REFERENCE
AUTHORS   11 (bases 1 to 303250)
          Takami, H. and Takaki, Y.
TITLE     Direct Submission
JOURNAL   Submitted (22-MAR-2000) Hideto Takami, Japan Marine Science and
          Technology Center, Deep-sea Microorganisms Research Group; 2-15

```



```

||||| ||||||| ||||| ||||| |||||
Db 241 TGTCTTATTGTAGTCAGACGATTGTCACCAATGAATTTATTGAAGAATAGATCAT 300
Qy 123 Glu 123
Db 301 AAA 303

RESULT 19
AL591983/c AL591983 324050 bp DNA linear BCT 06-JUN-2002
LOCUS Listeria monocytogenes strain EGD, complete genome segment 11/12.
DEFINITION AL591983 AL591824
ACCESSION AL591983.1 GI:16411809
VERSION AL591983.1 GI:16411809
KEYWORDS
SOURCE Listeria monocytogenes.
ORGANISM Listeria monocytogenes
Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
REFERENCE 1
AUTHORS
Glaser, P., Frangeul, L., Buchrieser, C., Rusniok, C., Amend, A.,
Baquero, F., Berche, P., Bloeker, H., Brandt, P., Chakraborty, T.,
Charbit, A., Chetouani, F., Couve, E., de Daruvar, A., Dehoux, P.,
Domann, E., Dominguez-Bernal, G., Duchaud, E., Durant, L.,
Dussurget, O., Entian, K. D., Fsihi, H., Portillo, F. G., Garrido, P.,
Gautier, L., Goebel, W., Gomez-Lopez, N., Hain, T., Hauf, J.,
Jackson, D., Jones, L. M., Kaerst, U., Kref, J., Kuhn, M., Kunst, F.,
Kurapkut, G., Madueno, E., Maitournam, A., Vicente, J. M., Ng, E.,
Nedjari, H., Nordsiek, G., Novella, S., de Pablos, B., Perez-Diaz, J. C.,
Purcell, R., Rammel, B., Rose, M., Schlueter, T., Simoes, N.,
Tierrez, A., Vazquez-Boland, J. A., Voss, H., Wehland, J. and Cossart, P.
Comparative genomics of Listeria species
Science 294 (5543), 849-852 (2001)
21537279
PUBMED 11679669
REFERENCE 2 (bases 1 to 324050)
AUTHORS
Glaser, P., Frangeul, L. and Rusniok, C.
Direct Submission
TITLE
Submitted (06-JUN-2001) Glaser P., Institut Pasteur, Genomique des
Microorganismes Pathogenes, 25 rue du Docteur Roux, 75724 Paris
JOURNAL
Cedex 15, FRANCE
COMMENT E-mail: pglaser@pasteur.fr
FEATURES
source
1. 324050
Location/Qualifiers
/organism="Listeria monocytogenes"
/strain="EGD-e"
/db_xref="taxon:1639"
complement(290..492)
/gene="lmo2321"
complement(290..478)
/gene="lmo2321"
/codon_start=1
/transl_table=11
/protein_id="CAD000399.1"
/db_xref="GI:16411810"
/db_xref="SPTREMBL:Q928W7"
/translation="MAERIPRKQTFIGNSEIFIDRTKMIANPAFRQKIALIETGCEK
MTDYIELKLKGYEEVTR"
complement(487..492)
/gene="lmo2321"
complement(586..801)
/gene="lmo2322"
complement(586..801)
/gene="lmo2322"
/codon_start=1
/transl_table=11
/product="gp44 [Bacteriophage A118]"
/protein_id="CAD00400.1"
/db_xref="GI:16411811"
/db_xref="SPTREMBL:Q8Y4V8"
/translation="MSEVLVSASYEGYKSKINSFTENNIVKRFKKIDVEKKRAE
TFNKKYEVTRKELVDGHLREIIIPRTL"
complement(join(798..1331,1340..1345))
/gene="lmo2323"

CDS
complement(798..1331)
/gene="lmo2323"
/codon_start=1
/transl_table=11
/product="gp43 [Bacteriophage A118]"
/protein_id="CAD00401.1"
/db_xref="GI:16411812"
/db_xref="SPTREMBL:Q8Y4V7"
/translation="MFQKSTSAAMQVLAETRTOKELATDSVYVTPALISNOIKKRT
VSLQAEQLIDSYNERSYLFYFAHFSNGMIPPLFDGLDNNHASLNNRFELEVEEAMN
TLKNGLETFTFLNRKGDMLQREAAKQAISETDVVASULTLNASIAIAFNIDLCQVLN
KRDLYYQKSGLVRSCKG"
complement(810..815)
/gene="lmo2322"
complement(810..815)
/gene="lmo2322"
complement(810..815)
/gene="lmo2322"
complement(1340..1345)
/gene="lmo2323"
complement(1454..2230)
/gene="lmo2324"
complement(1454..2230)
/gene="lmo2324"
/note="similar to anti-repressor [Bacteriophage A118]"
/codon_start=1
/transl_table=11
/protein_id="CAD00402.1"
/db_xref="GI:16411813"
/db_xref="SPTREMBL:Q8Y4V6"
/translation="MSNLOVIANEMLPVLNEOGEKFVDARMLHEKLLVNTFRNDWIY
RMIGNGYENGLDFYLSKTINGRPSLTNYFLTLDTAKELAMVQNNEMGRAIRKYFIEV
EKQARKLAYEPAFSYMIEDPVARAKKWIIEQOEVLKLEQPKPVVFAEAVQTS
ENTILVKDLATILKQKGLDIGONRLGFWLGRSGYLLSKGAYYKPKSQKAMNIGLFEQK
THIHDRNGLMKTTYTPQVTGKGQVYLLNKLLEHNQVII"
complement(2212..2487)
/gene="lmo2325"
complement(2212..2487)
/gene="lmo2325"
/codon_start=1
/transl_table=11
/protein_id="CAD00403.1"
/db_xref="GI:16411814"
/db_xref="SPTREMBL:Q8Y4V5"
/translation="MIFTVNSFPQSGHVFTPGLTVNTCEHSGOLPSPVFLIISCFAIF
RLCSLICLSNFFESNLILPILPTSLHKNYSTVKGRTERTKQIYK"
complement(2236..2241)
/gene="lmo2324"
complement(2236..2241)
/gene="lmo2324"
complement(join(2493..2774,2782..2787))
/gene="lmo2326"
complement(2493..2774)
/gene="lmo2326"
/note="similar to protein gp41 [Bacteriophage A118]"
/codon_start=1
/transl_table=11
/protein_id="CAD00404.1"
/db_xref="GI:16411815"
/db_xref="SPTREMBL:Q8Y4V4"
/translation="MNNIKQAIILKLETTILENGNAIESGSPVKYSVTKNLSLLEKQOE
LKIIEMKVELNGVEDSINATLLEKRLIEAKSLVEDLASTINSLEIKVK"
complement(2498..2503)
/gene="lmo2325"
complement(2498..2503)
/gene="lmo2325"
complement(2782..2787)
/gene="lmo2326"
complement(2800..3097)
/gene="lmo2327"
complement(2800..3084)
/gene="lmo2327"
/transl_table=11
/protein_id="CAD00405.1"

```

```
QY 75 ArgGlyPheTyrLysAspAlaSerThrIleArgLys-----AsnSerArgIleSer 91
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 87104 AAAGGTTCTTTGAAGACGACAAC-----CGTAAAGCCGCATACGACACAAGATTAGC 87051

QY 92 ArgLeuGluAspTyrIleLysGluTyrCysAsnPheGlyCysAlaTyrPheValLeuGlu 111
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 87050 ACCTTAAAGAATATATTTACGAACTCTTGCGTGTGCTTATTTTGTATT--- 86994

QY 112 AsnProAsnProArgAspIleLys----- 119
:::|||||
Db 86993 -----AAAAAGTAAATAGTAAATGAGGTAATTTGGAATGAAGAATTATAA 86946

QY 120 -----PheAspAspGluArgProHisLysArgArgLysSerArgSerLysSerGlnSer 137
||||:|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 86945 AGCGTATTAAATTGATTAGACGGCACAATGTATCGCGGAGCAGAAGTTATCCCGAAGC 86886

QY 138 Ser 138
:::
Db 86885 AAT 86883

RESULT 17
AX417048/c 349980 bp DNA linear PAT 14-JUN-2002
LOCUS
DEFINITION Sequence 4039 from Patent WO0228891.
ACCESSION AX417048
VERSION AX417048.1 GI:21449658
KEYWORDS
SOURCE Listeria innocua.
ORGANISM Listeria innocua
REFERENCE 1
AUTHORS Glaser, P. and Kunst, F.
TITLE Listeria innocua, genome and applications
JOURNAL Patent: WO 0228891-A 4039 11-APR-2002;
Pasteur Institut (FR)
FEATURES
Location/Qualifiers
1..349980
/organism="Listeria innocua"
/db_xref="taxon:1642"
/note="seq 2058, original length: 3,011,208 replaced
by-seq 2058: 0,000,001 to 0,349,980-seq 4032: 0,300,001 to
0,649,980-seq 4033: 0,600,001 to 0,949,980-seq 4034:
0,900,001 to 1,249,980-seq 4035: 1,200,001 to
1,549,980-seq 4036: 1,500,001 to 1,849,980-seq 4037:
1,800,001 to 2,149,980-seq 4038: 2,100,001 to
2,449,980-seq 4039: 2,400,001 to 2,749,980-seq 4040:
2,700,001 to 3,049,980-seq 4041: 3,000,001 to 3,011,208"
BASE COUNT 101702 a 72631 c 57863 g 117784 t
ORIGIN

Alignment Scores:
Pred. No.: 2e-11 Length: 349980
Score: 248.00 Matches: 63
Percent Similarity: 57.76% Conservative: 30
Best Local Similarity: 39.13% Mismatches: 35
Query Match: 24.58% Indels: 34
DB: 6 Gaps: 7

US-09-868-352-23 (1-192) x AX417048 (1-349980)

QY 2 ArgLysGluVal---ThrProGluMetLeu-----AsnTyrAsnLys 14
|||||
Db 124920 CGAAAGAAGTATATACACAGAAATTACTCCGAGATATTTAAATGGAACTCAAAAA 124861

QY 15 TyrProGlyProGlnPheIleHisPheGluAsnIleValLysSerAspIleGluPhe 34
|||:::|
Db 124860 ATGTACGCACAGGAGGA-AGACACGCTG-----ACGATTACGATTCAAGATTAAACTAC 124808

QY 35 GlnLeuValIleAsnGlnLysSerAlaPheAspValThrValPheGlyGlnArgPheSer 54
:::|||||
Db 124807 GAAATTATCACAATTATCCGAGTCTTTTGACGAGGAAAAATTGAATGAACGATTTAGC 124748

QY 55 GlulleLeuLeuLysTyrAspPheIleValGlyAspTrpGlyAsnGluGlnLeuArgLeu 74
```

```
Db 124747 GATATTCTTGGCGGATGATGATTATATAGTAGCGACTGGGTTACGATCAACTTAGACTG 124688

QY 75 ArgGlyPheTyrLysAspAlaSerThrIleArgLys-----AsnSerArgIleSer 91
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 124687 AAAGGTTCTTTGAAGACGACAAC-----CGTAAAGCCGCATACGACACAAGATTAGC 124634

QY 92 ArgLeuGluAspTyrIleLysGluTyrCysAsnPheGlyCysAlaTyrPheValLeuGlu 111
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 124633 ACCTTAAAGAATATATTTACGAACTCTTGCGTGTGCTTATTTTGTATT--- 124577

QY 112 AsnProAsnProArgAspIleLys----- 119
:::|||||
Db 124576 -----AAAAAGTAAATAGTAAATGAGGTAATTTGGAATGAAGAATTATAA 124529

QY 120 -----PheAspAspGluArgProHisLysArgArgLysSerArgSerLysSerGlnSer 137
||||:|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 124528 AGCGTATTAAATTGATTAGACGGCACAATGTATCGCGGAGCAGAAGTTATCCCGAAGC 124469

QY 138 Ser 138
:::
Db 124468 AAT 124466

RESULT 18
AX142599
LOCUS
DEFINITION Sequence 1321 from Patent WO0134809.
ACCESSION AX142599
VERSION AX142599.1 GI:14282171
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct
REFERENCE 1
AUTHORS Kimmerly, W.J.
TITLE Staphylococcus epidermidis nucleic acids and proteins
JOURNAL Patent: WO 0134809-A 1321 17-MAY-2001;
GLAXO GROUP LIMITED (GB)
FEATURES
Location/Qualifiers
1..390
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="synthetic nucleic acid sequence"
BASE COUNT 158 a 41 c 67 g 124 t
ORIGIN

Alignment Scores:
Pred. No.: 9.32e-15 Length: 390
Score: 243.50 Matches: 45
Percent Similarity: 68.32% Conservative: 24
Best Local Similarity: 44.55% Mismatches: 29
Query Match: 24.13% Indels: 3
DB: 6 Gaps: 2

US-09-868-352-23 (1-192) x AX142599 (1-390)

QY 26 IleValLysSerAspIleGluPheGlnLeuValIleAsnGluLysSerAlaPheAsp 45
:::|||||
Db 1 ATGATTAAAGTCGACCAACAATATTTTGAATTGATAGAGAATATAGAGAATGTTTGTAT 60

QY 46 ValThrValPheGlyGlnArgPheSerGluIleLeuLeuLysTyrAspPheIleValGly 65
:::|||||
Db 61 GAGGAATATTTTACGCTAGGTATTCGGATATATTCGGAATATATGACAAATATGATTATCGTAGGT 120

QY 66 AspTrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAspAlaSerThrIleArg 85
|||:::|
Db 121 GACTATGTTACCATCAATTTACCTTAAAGAGATTTTATAAAGTAGTATAAAGGCCA 180

QY 86 Lys---AsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyrCysAsnPheGly 104
:::|||||
Db 181 GAAATAAGTAACGATTTTCAAGTATACAAGATTATATATACTAGAATATTCGTAATTTTGGT 240

QY 105 CysAlaTyrPheValLeuGluAsnProAsnProArgAspIle-----LysPheAspAsp 122
```

```
/transl_table=11
/protein_id="CAC97562.1"
/db_xref="GI:16414858"
/db_xref="SPTREMBL:Q929E2"
/transl_table="MNJHURFYVTLAHMEHYTKAAEKLLITOPSLSHAITSLEQELG
IPLEKGRNIGLSKAGKGVVYVEESLMDIDAGVATTKKAANGEGQIDLAFLQTLGT
SLVKLVQVEIQTEPEKKIDVFHTGVSIDIIGLEKKKYDIGICSLERNRHFPTP
IAKELVLIYVPKSPHAEKDWIDLIETVPYPIAFSKKSLRPIIDDLERKIGDYKI
AYETEVDVITAGWAQNGFATVPNMFILNVVDVKVLPISPNWERFEYLAVYKNQL
TPAERKQWLEHRI"
8543. .9397
/gene="lin2336"
CDS
8543. .9397
/gene="lin2336"
/feature="similar to unknown proteins"
/codon_start=1
/transl_table=11
/protein_id="CAC97563.1"
/db_xref="GI:16414859"
/db_xref="SPTREMBL:Q929E1"
/transl_table="MTNAGDLKCPITISSYTLGTEVSPERVRIAAENGFDGIGLR
AENVYDALAGLTDDEPRLIDENIKVTEYITQMGTAEDRTVEQKKKEOTFFHMA
RLFVGHINGCLEKIPERQIITALGELCDRAEELITGLEFMPYSGVADLAAARVAE
ACGRDNLICDTHHARNOQTESIKNPADRIYSIQLCDVHETPYKELRESLHDKR
LAPEGYGDYVGFARILKEHGVNPRVMEVSDSMVETGLEVAAIKVYNATKKVLDE
AWPEISPK"
9467. .11392
/gene="lin2337"
CDS
9467. .11392
/gene="lin2337"
/feature="similar to NADH oxidase"
/codon_start=1
/transl_table=11
/protein_id="CAC97564.1"
/db_xref="GI:16414860"
/db_xref="SPTREMBL:Q929E0"
/transl_table="MKFNSMFSPIDIGMRVFNRFVSPMCMNYANTDGLTDTSLAY
YKRALGFGFLITEFVTVDRAGKAGKACLYSDHOIASFKRVIDACHDAGAKISVO
LOHAGPENSKVSQYPLKASATASAEGRNTPPAISREELYELIELYGEALRAKKAG
ADAVEICANGLYSVSLSGRTNKRVDGCGCFENRMRPLRIIESIRKRVGHISAIL
CRINTQDGLSVQSDATVAAYLEDCGLDGLHVSXVHIRDEYMWAPITVLHAGFSS
DLNTOIKRAVSIPIVTRTEPHYAEMLVREGADLVAFGRSLADPPTPNKAAAGK
IDELLPCACLOGVANNYAGKPTICLVNPLLGRESEAYLPKTPSKVYVIGGVGL
YAGMAGSRGHDVTVYEAASDIIGOMRLAAYPGKGLDITNNVRSYIKKQBFQVEIKT
NTPVTPELIODIAPDAVIATATGATPLVLPIDGSLIHAVDLDDCKESCQKVLVY
GGMVGSETAFLGEAGHDYTVVELRDEVGADVISEHRKFLMDFDFEYIKSTINAKV
TSFLPDGVTSLADEKRYIDGDFSVVLAMGSRAYNPLEEAIKRVIPETIVIGDAVRA
RRALDATKEALDAVLQ"
```

Alignment Scores:

Pred. No.:	1.29e-11	Length:	248050
Score:	248.00	Matches:	63
Percent Similarity:	57.76%	Conservative:	30
Best Local Similarity:	39.13%	Mismatches:	35
Query Match:	24.58%	Indels:	34
DB:	1	Gaps:	7

US-09-868-352-23 (1-192) x AL596172 (1-248050)

Qy 2 ArgLysGluVal---ThrProGluMetLeu-----AsnTyrAsnLys 14
|||||
Db 162920 CGAAAAGAAGTATATACACCAAGATTACTCCGAGATATTTTAAATGGAACCTCAAAAAA 162861

Qy 15 TyrProGlyProGlnPheIleHisPheGluAsnIleValLysSerAspIleGluPhe 34
|||||
Db 162860 ATGTACGACAGAGAGGA-AGACACGTG-----ACGATTACGATTCAAGATTAACTAC 162808

Qy 35 GlnLeuValIleAsnGluLysSerAlaPheAspValThrValPheGlyGlnArgPheSer 54
|||||
Db 162807 GAAATTATCACCATTATTCGGATGCTTTTACCGAGGAAAAATTTGAATGAACGATTATAGC 162748

Qy 55 GluIleLeuLeuLysTyrAspPheIleValGlyAspTrpGlyAsnGluInLeuArgLeu 74
|||||

```
Db 162747 GATATTTCTGGCGGATATGATTATATAGTAGCGACTGGGCTTACGATCAACTTAGACTG 162688

Qy 75 ArgGlyPheTyrLysAspAlaSerThrIleArgLys-----AsnSerArgIleSer 91  
|||||  
Db 162687 AAAGGTTTCTTTGAAGACGACAAC-----CGTAAAGCCGATACGACAACAAGATTAGC 162634



Qy 92 ArgLeuGluAspTyrIleLysGluTyrCysAsnPheGlyCysAlaTyrPheValLeuGlu 111  
|||||  
Db 162633 ACGTTAAAGAATATATTTAGCACTACTGCAACTTGGCTGTGCTTATTTGTTATT--- 162577



Qy 112 AsnProAsnProArgAspIleLys----- 119  
|||||  
Db 162576 -----AAAAAGTAAATAGTAAATAGGTAATTTGGAATTGAAGAAATTATAA 162529



Qy 120 -----PheAspAspGluArgProHisLysArgArgLysSerArgSerLysSerGlnSer 137  
|||||  
Db 162528 AGCGTATTTAATGATTAGACGCGCACAAATGATCGCGGACGAGAAGTTATTCGCGAAGC 162469



Qy 138 Ser 138  
|||



Db 162468 AAT 162466



RESULT 16  
AX413018/c 349980 bp DNA linear PAT 14-JUN-2002  
LOCUS AX413018  
DEFINITION Sequence 9 from Patent WO0228891.  
ACCESSION AX413018  
VERSION AX413018.1 GI:21445476  
KEYWORDS  
SOURCE Listeria innocua.  
ORGANISM Listeria innocua  
REFERENCE 1  
AUTHORS Glaser, P. and Kunst, F.  
TITLE Listeria innocua, genome and applications  
JOURNAL Patent: WO 0228891-A 9 11-APR-2002;  
Pasteur Institut (FR)  
FEATURES  
source  
1. 349980  
/organism="Listeria innocua"  
/db_xref="taxon:1642"  
/note="old seq 9 (684707) splitted in 3-new seq 9 = old  
seq 9 from 1 to 349980-seq 4027 = old seq 9 from 300000 to  
649980-seq 4028 = old seq 9 from 600000 to 684707"



BASE COUNT 104581 a 69113 c 62414 g 113872 t



ORIGIN



Alignment Scores:



|                        |        |               |        |
|------------------------|--------|---------------|--------|
| Pred. No.:             | 2e-11  | Length:       | 349980 |
| Score:                 | 248.00 | Matches:      | 63     |
| Percent Similarity:    | 57.76% | Conservative: | 30     |
| Best Local Similarity: | 39.13% | Mismatches:   | 35     |
| Query Match:           | 24.58% | Indels:       | 34     |
| DB:                    | 6      | Gaps:         | 7      |



US-09-868-352-23 (1-192) x AX413018 (1-349980)



Qy 2 ArgLysGluVal---ThrProGluMetLeu-----AsnTyrAsnLys 14  
|||||  
Db 87337 CGAAAAGAAGTATATACACCAAGATTACTCCGAGATATTTTAAATGGAACCTCAAAAAA 87278



Qy 15 TyrProGlyProGlnPheIleHisPheGluAsnIleValLysSerAspIleGluPhe 34  
|||||  
Db 87277 ATGTACGACAGAGAGGA-AGACACGTG-----ACGATTACGATTCAAGATTAACTAC 87225



Qy 35 GlnLeuValIleAsnGluLysSerAlaPheAspValThrValPheGlyGlnArgPheSer 54  
|||||  
Db 87224 GAAATTATCACCATTATTCGGATGCTTTTACCGAGGAAAAATTTGAATGAACGATTATAGC 87165



Qy 55 GluIleLeuLeuLysTyrAspPheIleValGlyAspTrpGlyAsnGluInLeuArgLeu 74  
|||||



Db 87164 GATATTTCTGGCGGATATGATTATATAGTAGCGACTGGGCTTACGATCAACTTAGACTG 87105


```

```

ORGANISM      Listeria innocua
REFERENCE      1
AUTHORS        Glaser,P., Frangeul,L., Buchrieser,C., Rusniok,C., Amend,A.,
                Baquero,F., Berche,P., Blocker,H., Brandt,P., Chakraborty,T.,
                Charbit,A., Chetouani,F., Couve,E., de Daruvar,A., Dehoux,P.,
                Domann,E., Dominguez-Bernal,G., Duchaud,E., Durant,L.,
                Dussurget,O., Entian,K.D., Fsihi,H., Portillo,F.G., Garrido,P.,
                Gautier,L., Goebel,W., Gomez-Lopez,N., Hain,T., Hauf,J.,
                Jackson,D., Jones,L.M., Kaerst,U., KrefT,J., Kuhn,M., Kunst,F.,
                Kurapat,G., Madieno,E., Maitournam,A., Vicente,J.M., Ng,E.,
                Nedjari,H., Nordsiek,G., Novella,S., de Pablos B., Perez-Diaz,J.C.,
                Purcell,R., Rammel,B., Rose,M., Schlueter,T., Simoes,N.,
                Tierrez,A., Vazquez-Boland,J.A., Voss,H., Wehland,J. and Cossart,P.
                Comparative genomics of Listeria species
                Science 294 (5543), 849-852 (2001)
                21537279
                PUBMED 11679669
REFERENCE      2 (bases 1 to 248050)
AUTHORS        Glaser,P., Frangeul,L. and Rusniok,C.
TITLE          Direct Submission
JOURNAL        Submitted (09-JUL-2001) Glaser P., Institut Pasteur, Genomique des
                Microorganismes Pathogenes, 25 rue du Docteur Roux, 75724 Paris
                Cedex 15, FRANCE
COMMENT        E-mail: pglaser@pasteur.fr
                Phone: +33 (0)1 45 68 89 96, Fax: +33 (0)1 45 68 87 86.
FEATURES       Location/Qualifiers
                1..248050
                /organism="Listeria innocua"
                /strain="Clip11262"
                /db_xref="taxon:1642"
gene            complement(join(731..1633,1644..1649))
                /gene="lin2330"
CDS             complement(731..1633)
                /gene="lin2330"
                /note="similar to ABC transporter (ATP-binding protein)"
                /codon_start=1
                /transl_table=11
                /protein_id="CAC97557.1"
                /db_xref="GI:16414853"
                /db_xref="SPTREMBL:Q929E7"
                /translation="MKLEIOLGKKYCKWIKIGDITNNKECVALIGPNACKST
                MNIVQILAQDEGOIKLQDQAKLVREKIGELPOYPPEYGMMSALECLRFMCKLSNM
                KKOOLEERKEVLLVGLADSTNKKISYSGMRGLGIAQAILHRPELLVDEPVS
                LDPIGRREIVLEKLSKEMTLFSTHLKDAEEICDRIATIKNGELVADKTVEOLMQ
                EESSPVDFEFTGDEAWLALNQNKDIEKIEKVGWVQVYVAYAIKKGAGSEATLEALK
                TKGELIRVENHQSLSEIFMEKVR"
                complement(739..744)
gene            /gene="lin2329"
                complement(739..744)
RBS             /gene="lin2329"
RBS             /gene="lin2329"
                complement(1644..1649)
RBS             /gene="lin2330"
gene            complement(11898..4091)
                /gene="lin2331"
                complement(1898..1929)
                /gene="lin2331"
                complement(1936..4080)
                /gene="lin2331"
                /note="similar to penicillin-binding protein"
                /codon_start=1
                /transl_table=11
                /protein_id="CAC97558.1"
                /db_xref="GI:16414854"
                /db_xref="SPTREMBL:Q929E6"
                /translation="MDKFKQOLIKVYKSFEGFIGKWIWLBKFRFEMKNKHGKIFL
                LAGLVFLSFYIYLVVAKSADIALKKGLESATIVYDKDGKAGELSTDTQVSD
                LKSNLQNAVSVIEDRRFYEHKGFDLKGARAGVNLITNGSGGGSTITQQLKNA
                LTKNQFTTKRAEIFEMAREIEKYSKDEIMENLNRSEYNGEWENASLKYFGKPA
                ADLNPTEAATIAGLQAPSAYPDYPQIHIEKATNRRNMLNAMVETGDIKAEQDKYKAT
                KIVLNDOSKDLANKYIPWVDVAVINEAVNEADITODEIMOKGYKIYTELDQNVQTSLE
                KYVNDLSLPRANDGTLVQSGAVLMDPATGGIRALVGRGHEHVFGRNRTQMKRAQP
                GSTMKPLAVITFALQSGYNVDSMLKDEKVTYKGNTPPTNVGGIYNGEVPYKAVANSI

```

```

RBS             complement(4086..4091)
gene            /gene="lin2331"
                4388..4860
                /gene="lin2332"
RBS             4388..4394
                /gene="lin2332"
CDS             4404..4829
                /gene="lin2332"
                /note="similar to arsenate reductase"
                /codon_start=1
                /transl_table=11
                /protein_id="CAC97559.1"
                /db_xref="GI:16414855"
                /db_xref="SPTREMBL:Q929E5"
                /translation="MTQKLIYFLSQTHIRSAIAEAWAKKLSLSNVKFIISGSHWHSKAT
                PFIAELNEFAIEPPESISYSPSELLADADLIVTIDYSAHETAPQFPANIOEKIY
                DIDDPEQETALPKWASQEVCDNIALSVKNLEHVLIEA"
                4838..4860
                /gene="lin2332"
                4877..5761
gene            /gene="lin2333"
                4877..4882
RBS             /gene="lin2333"
                4892..5761
CDS             /gene="lin2333"
                /note="similar to unknown proteins"
                /codon_start=1
                /transl_table=11
                /protein_id="CAC97560.1"
                /db_xref="GI:16414856"
                /db_xref="SPTREMBL:Q929E4"
                /translation="MDGYNDLKAKEAFLSIFAYLFLSVLKIAGHFNSDALLADG
                LNHTTIDIVASVALLIGLRISRIPPDADSHSYGHRRTETSSLIASIIIMFLVGQVWSS
                LVHIIKEKETAPSMLETAIVALESGVFMAYIYVNHRLAKKLDLSQAVRAAYDNRSDF
                VSGFAGIGIGAVLGVPLDVSITAFVLGVLYIVTAVKIFYDAAHTLTDGFDVSKLETI
                HDLIASVPDVKKVIDIKARMNGNRWIDATIAVDPLNVVKSHEITEIVEQIRNEYE
                GAFTLVHLEPFE"
gene            join(6009..7323,7331..7360)
                /gene="lin2334"
RBS             6009..6014
                /gene="lin2334"
                6019..7323
CDS             /gene="lin2334"
                /note="similar to unknown proteins"
                /codon_start=1
                /transl_table=11
                /protein_id="CAC97561.1"
                /db_xref="GI:16414857"
                /db_xref="SPTREMBL:Q929E3"
                /translation="MILTIFLIILAIASFAFFVATEFAIVKMRPSRLDQLIAEKDK
                RAVLARHIYNHLNAYLSACQLGIISSGLGWLGEVTEAALHPLFSLLEPQSATTI
                LSFTTAFLEFTHVVGELVPKLAIDTETVALAVARPLHIFYKMPFPFWILNGS
                AVFIARFLGLEPASEHEIAHTDELKLIIVGESYKGEINOSEFRVYVKNLKFDFDERMAK
                EVMIPRTEIVTDTGTIGELSDIMONERYTRYPVIDGDKDHVIGVNLNLKELSAVVE
                HGSNPSFSDIPYVKPIIRVIEITPIKELFLRQWRSHIAILLDEVGSTSGLVTVEDI
                VEEIVDIRDEFDADEPIEIRIKGDHYIVDAKLLIDEVNNILGTETEEEEEDVTIGGW
                FLTQNYEVEVGDEIDYDGFIFRVKQGEQHIEYIEIIRKNTD"
                complement(join(7331..7360,7364..8242))
                /gene="lin2335"
                7331..7360
gene            /gene="lin2334"
                complement(7331..7360)
terminator      /gene="lin2335"
                complement(7364..8242)
CDS             /gene="lin2335"
                /note="similar to transcriptional regulators (LysR
                family)"
                /codon_start=1

```


gene	<pre>/gene="ilvE" /notes="ORFID:SA0512 branched-chain amino acid aminotransferase homologue" /codon_start=1 /transl_table=11 /protein_id="BAB41743.1" /db_xref="GI:13700446" /translation="MSQAVRVERETLKQRPNTSQLGFGKYFTDYMLSYDYDADKGWH DKIVPGPTEISPAQGVHVGQSVEFGLKAYKRDGEVALFRPENEKRLNNSLARLE MPQVDEALELGLKQVLDIERDMQIPEGEGOSLXIRPEVEFEGALGVGASHOYKLLII LSPSAYIGGETLKPTKIYVEDEVYRAVRGGVGFAGVAGNAAASLLAQTNANKLGYDQ VLMDGVEQYKIEVGSNMIFVENGKVIITPELNGSLIPGTRKSIIEIAKNGUYEVE ERRYSIDELFESYDKBELTEFVFGSGTAAVISPVGTLRYEDREIVINNETGEITQKLY DVTGIGQNGTLEDKNGRWVVVPKY" 1574. .2257 /gene="SA0513" 1574. .2257 /gene="SA0513" /notes="ORFID:SA0513" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="BAB41744.1" /db_xref="GI:13700447" /translation="MEWILFDKGTGLIEFDRSWEKIGRVFVQSILLETFFVHNKEAALR QLGVIKESIDPKVMGSGSLQIIQAFNDVTGQDITDWSKSTOKLVDRIPEINWVE GVKEALDLKAKGVOLGIVTSDTKKGVEQFLAHTNALSLFDLIIISTEADAYEKPNKV LSPLEQYVDPQKVAIVGDTANDMTASNANLGMALGVLGTGTATKEELHEADLIINS AADILEALN" complement(2373. .3035) /gene="SA0514" complement(2373. .3035) /gene="SA0514" /notes="ORFID:SA0514 hypothetical protein, similar to deoxypurine kinase" /codon_start=1 /transl_table=11 /protein_id="BAB41745.1" /db_xref="GI:13700448" /translation="MNNYGIPQNAITIACTVGVGKSTLTQALADKLNKFTSFENVEH NPYLDKFSYDERWSFHLQIYFLAERFKQKRFMEYGGGVQDRSIVYEDVIFAKMHE EEGTSKEDFKTYSDFNAMVMTYPYKPDVMYILECNYDEVIDRIIERGREMEINTD PEYMKLFKRYDDWINSFNACPVVRINNEYDIHKDPESLNPMDIKIARIQIYQVD TR" complement(3028. .3645) /gene="SA0515" complement(3028. .3645) /gene="SA0515" /notes="ORFID:SA0515 hypothetical protein, similar to deoxypurine kinase" /codon_start=1 /transl_table=11 /protein_id="BAB41746.1" /db_xref="GI:13700449" /translation="MKNPFTAIIEGPIGVGKSLAHLSQTLDFYEKEIITEPNPFLSD FYEDIKSWSTQEMFFLCNRYKQFDVTLNQGVSDYHIIHKNIIFAKNPLSPVEFOK FSKIYDILTDMTPNMNIIIFLDADLDVLKRIAKRNSFEHQIEDEYLLKDKDYREY YESIQSGNSNVLLDITSDIFLNEQDYEDILHILPIMGDIENE" 3712. .4182 /gene="SA0516" 3712. .4182 /gene="SA0516" /notes="ORFID:SA0516 hypothetical protein, similar to Cu binding protein (Mn oxidation)" /codon_start=1 /transl_table=11 /protein_id="BAB41747.1" /db_xref="GI:13700450" /translation="MNDIYFMTLATEEAKKAQGEVPIGAIITKDDEVIARAHNLR ETIQQTATAHEHTAIEAAKVLGSRLEGCTLTLTLEPCVMCAGTIVMSKIPRVYGA DDPKGCGSGSLMNLQOSNFNHRAIYDKGVLRKACSTLLTTFEKNLRANKKSTN" 4329. .5198</pre>	CDS	<pre>/gene="SA0517" 4329. .5198 /gene="SA0517" /notes="ORFID:SA0517" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="BAB41748.1" /db_xref="GI:13700451" /translation="MIKLIATDMGDTLLNAAHEISQPNIDAIKYAOEQGITVVIATGR AFTEAQPAPVADTLTVPICLNGAEVRDETFNVMSTSHLKNLSLVHKLTINVLKDGIIY ENPISRAYITEDFQRLDIYDIAERAQGHANVERIKNGIQRRIIDNTGLKVDYDAI ENIPGELIMKILAFDGNLEKILAKESPNLAISSSRGNIEITHSDAQKIGALE TIAERLGIEMKDVMAIGDNLNLSMLEKVGYPVAMENGAEVKKIAKYVTDITNENSGV GKALMKLLREQV" 5220. .5786 /gene="SA0518" 5220. .5786 /gene="SA0518" /notes="ORFID:SA0518" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="BAB41749.1" /db_xref="GI:13700452" /translation="MKGLIITIGSAQVNSHTSALARYLTHEFKTHDIEAEIFDLAEKPL NQIDFSGTTFSEIDEIKQNMKDLAEKAMAADFLILGTPNYHGSYSGILKNALDHLNMDY FRMKPGLIGNSGGIVSSSEPLSHLRVIVRSLIGLIAVPTQIATHDSDFAKNEDGSIYLN DSEFQLRARLFDQIVSFVNNSPYEHLK" 6215. .9076 /gene="sdrC" 6215. .9076 /gene="sdrC" /notes="ORFID:SA0519" /codon_start=1 /transl_table=11 /product="Ser-Asp rich fibrinogen-binding, bone sialoprotein-binding protein" /protein_id="BAB41750.1" /db_xref="GI:13700453" /translation="MNNKTATNRKGMIPNRLNKFISIRKYSGVTASILVGTTLIFGLS GHEAKAEHTNGELNQSNETTAPSENKTKTEYDSQKLDNTQTATADQPKVMSDSA TVKETSSMOSQPNATASQSTQTQTSNVTNDKSTTYSNETDKNTLQAQKNVSTTPKT TTIKQALNRMAVNTAAPOQGTNVNDKVHFTNIDIAIDKHVNTKTTGNTFEWATSSD VLKAKANYTIDDSVKEDDTETFKYGOYFRPGSVRLPSQTONLYNAOGNIIKAGLYDSK TNTTTFYTNVQYTNVSSGFEQVAFKRENATDTKATKMEVTLGNDIYSKDVIYD YGNQKQGLISSTNYINNEDLSRNMTYVNPQKTYTKTFVTNLTKYKFNPAKNFK IYEVTQNGFVDSFTPTSKLDVTGQFDVIYSDNKTATVOLLGSSSDSKQYIIQQ VAYPDNSSTDNGIDYTLQNGKSSNSYSNVNGSSSTANGQKQKYNLGDYVWEDTN KDGQDANERKIGKGVILKDSNGKELDRITTDENGKYQFTGLSNGYSVEFSTPAGY TPTTANAGTDADVDSGLTGTGVIKADANNLTDSGFYKTPKYSIGDYVWVDSNKGDKQ DSTEKGIKGVKPTLQNEKGEVIGTTEDENGKYRFNLDGSKYKVIFEKPAGLTQGT NTTEDKADGGEVDVTITDHDFTLNGYIEETSDSDSDSDSDSDSDSDSDSDSDSDSD SD SD SD ALPETGSENNNSNNGTLFGCLFAALGSLLLFGRRKKQNK" 9443. .13600 /gene="sdrD" 9443. .13600 /gene="sdrD" /notes="ORFID:SA0520" /codon_start=1 /transl_table=11 /product="Ser-Asp rich fibrinogen-binding, bone sialoprotein-binding protein" /protein_id="BAB41751.1" /db_xref="GI:13700454" /translation="MLNRENKTAIRKGMVSNRLNKFISIRKYSGVTASILVGTTLIFG LGNQEKAAESTNKNELNEATTASDQSSQSDKQVDMQQLNQEDNTKDNQKEMVSSQNE TYSNGAKSEKESVQCTTGNKVEVSTADEQSPKSTNEDLNTKOTISNOEQLQDPL LENKSVYVQPTNEENKVKADKTESTTLAVKSDAIKSNAEITLVLDNNSNSNENNAADI LPKSTAPKSLNTRMAAIQPNSTDSKNVNDLITSNTTLTVVDADNSKTIIVPAQDYL</pre>	gene
------	--	-----	---	------

```

/protein_id="BAB94399.1"
/db_xref="GI:21203700"
/translation="MIIYRQVHNEGAPVEIITKTFQHSVKICDDSFSDTEIFKLLSL
LQDDIDHMKVS"
complement(6293..7123)
/gene="MW0535"
complement(6293..7123)
/gene="MW0535"
/note="ORFID:MW0535
hypothetical protein, similar to phosphomethylpyrimidine
kinase, thid homologue"
/codon_start=1
/transl_table=11
/protein_id="BAB94400.1"
/db_xref="GI:21203701"
/translation="MALKKVLTIAGSDTSAGAGMQADLKTFOELDTYGMVALTAVTM
DKDTWSDHVTPLPMDVDFKQLETALSIGPDAIKTGMGTETIIKRAGEVYEASNAQYF
VVDPMVCKGEDEVLPNGTEAMIKYLLPKATVTPNLFEGQLSLGKLNSIDPMKK
AATIFDGAQHVIIKKGKALDQDKSYDLVDGQTFYQLTDMFQGSNHHGACCTFAA
ATTYLDANGSKPEAVISAKAFVASAIRKNGMKMDFVGPVDHGAYNRIEHDIVVEVTE
"
7307..7963
/gene="ung"
7307..7963
/gene="ung"
/note="ORFID:MW0536"
/codon_start=1
/transl_table=11
/product="uracil-DNA glycosylase"
/protein_id="BAB94401.1"
/db_xref="GI:21203702"
/translation="MENSQIFHDITTKHDFKAMHDFLEKEYSTAIVYDPRENIYAQFD
LTPFNKIVVLGQDPYHPNGAHLAFSVQPNAPKPPSLRMNYKELADDIGCVRQTP
HQQWARGVELLNTVLTVRQGEANSHRDIGHWETFDIIRKAVSDYKHHVPIILWKP
AQQKIKLDTSKHCIIKSVHPSPLSAVRGFGSKPYSKANAYLESVGKSPINWCESEA
"
7964..8344
/gene="MW0537"
7964..8344
/gene="MW0537"
/note="ORFID:MW0537"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB94402.1"
/db_xref="GI:21203703"
/translation="MLNRETLIARIEQELVQAEQAQHDHDFEKHYAIHILTSYAST
SNTPHIGEQMMNRRIANHMOPSOITPTHQVTAETAEMGGKYNTHSAHHNKSYS
OPSNQORLATDDDIGNGESIFDF"
8477..8845
/gene="MW0538"
8477..8845
/gene="MW0538"
/note="ORFID:MW0538"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB94403.1"
/db_xref="GI:21203704"
/translation="MKLFILGALNMAVGTGAFGAHGLOGKIKSDHLYSLVWEKATTY
OMYHGLALLIGVIGSTSTINYNWAGLIFAGIIFSGSLYLTLVLTQIKVLGATPIG
GVLFILGIMLLIATFKFAG"
8966..10450
/gene="MW0539"
8966..10450
/gene="MW0539"
/note="ORFID:MW0539"

```

Alignment Scores:

Pred. No.:	1.13e-11	Length:	290150
Score:	249.50	Matches:	50
Percent Similarity:	62.28%	Conservative:	21
Best Local Similarity:	43.86%	Mismatches:	30

```

Query Match: 24.73% Indels: 13
DB: 1 Gaps: 2
US-09-868-352-23 (1-192) x AP004824 (1-290150)
QY 24 GluAsnIleValLysSerAspIleGluPheGlnLeuValIleAsnGluLysSerAla 43
:::||||:||||| ||| |||:||||: ||| |||
Db 279243 AAGAATTTGATAAAAGTAGATCAACATTACTTTGAATTAATAGAAATTTATCCGGAATGT 279302
:::||||:||||| ||| |||:||||: ||| |||
QY 44 PheAspValThrValPheGlyGlnArgPheSerGluIleLeuLeuLysTyrrAspPheIle 63
|||:||||:||||| ||| |||:||||: ||| |||
Db 279303 TTTAATGAAGAACAAATTTATTCTAGTAGTATTCAGATATTTAGATAAATATGATTACATA 279362
|||:||||:||||| ||| |||:||||: ||| |||
QY 64 ValGlyAspTrpGlyAsnGluGlnArgLeuArgGlyPheTyrrLysAspAlaSer--- 82
|||:||||:||||| ||| |||:||||: ||| |||
Db 279363 GTTGGTGACTAGTTTACGATCAATTACGATTAAAGAGTTTTTACAAAGATTTCTAATAAA 279422
:::||||:||||| ||| |||:||||: ||| |||
QY 83 ThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrrIleLysGluTyrrCysAsn 102
:::||||:||||| ||| |||:||||: ||| |||
Db 279423 AAAGCAGAGATGAGTAAAGCTTTTTCAAATATTCAAGATTACATATTTGAATATTGTAAC 279482
|||:||||:||||| ||| |||:||||: ||| |||
QY 103 PheGlyCysAlaTyrrPheValLeu----- 110
|||:||||:||||| ||| |||:||||: ||| |||
Db 279483 TTTGGTTGCTTACTTTGTTTAAACACATTTCTTAACAACAGGTTAAAGATTAAATC 279542
|||:||||:||||| ||| |||:||||: ||| |||
QY 111 GluAsnProAsnProArgAspIleLysPheAspGluArg 124
|||:||||:||||| ||| |||:||||: ||| |||
Db 279543 GAAGAAGTTTCACTCCGTCTGATGTAGATGACGACAATAAA 279584
|||:||||:||||| ||| |||:||||: ||| |||
RESULT 13
AP003131 299050 bp DNA linear BCT 02-JUL-2002
LOCUS Staphylococcus aureus subsp. aureus N315 genomic DNA, complete
DEFINITION genome, section 3/10.
ACCESSION AP003131 BA000018
VERSION AP003131.2 GI:14349173
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
Kuroda, M., Ohta, T., Uchiyama, I., Baba, T., Yuzawa, H., Kobayashi, I.,
Cui, L., Oguchi, A., Aoki, K., Negai, Y., Lian, J., Ito, T., Kanamori, M.,
Matsumaru, H., Maruyama, A., Murakami, H., Hosoyama, A.,
Mizutani-U, Y., Takahashi, N. K., Sawano, T., Inoue, R., Kaito, C.,
Sekimizu, K., Hirakawa, H., Kuhara, S., Goto, S., Yabuzaki, J.,
Kanehisa, M., Yamashita, A., Oshima, K., Furuya, K., Yoshino, C.,
Shiba, T., Hattori, M., Ogasawara, N., Hayashi, H. and Hiramatsu, K.
Whole genome sequencing of methicillin-resistant Staphylococcus
aureus
JOURNAL Lancet 357 (9264), 1225-1240 (2001)
MEDLINE 21311952
PUBMED 11418146
REFERENCE
AUTHORS Aoki, K., Oguchi, A., Hosoyama, A., Nagai, Y., Kuroda, M., Hiramatsu, K.
and Kikuchi, H.
Direct Submission
JOURNAL Submitted (30-JAN-2001) Akio Oguchi, National Institute of
Technology and Evaluation, Biotechnology Center; 2Chome 45-10
Nishihara, Shibuya-ku, Tokyo 151-0066, Japan
(E-mail:oguchi@nitech.go.jp, URL:http://www.bio.nite.go.jp/,
Tel:81-3-3481-8423, Fax:81-3-3481-8424)
On Jun 12, 2001 this sequence version replaced gi:13700445.
FEATURES
Location/Qualifiers
1..299050
/organism="Staphylococcus aureus subsp. aureus N315"
/strain="N315"
/sub_species="aureus N315"
/db_xref="taxon:158879"
248..1324
/gene="ilvE"
248..1324
CDS

```

[illegible]

Db	7324	--TCAACCAGTCGTAGAAAAATGTTGGCGACACAATAAAAATTGATCATCTTTGAAGATTAT			: : : : : : : :		7380
Qy	97	IleLysArgCysAsnPhgIeGlyCysAlaTyrrPheValLeuGluAsnProAsnProArg			: : : : : : : :		116
Db	7381	GTCAAATGAATATTGTAATTCAGGTTGTGCTTATTTTGTCTTTACGCCGGATTCCGCACAAA	:	:	: : : : : : : :	:	7440
Qy	117	AspIleLysPheAspGluArg-			: : : : : : : :		124
Db	7441	GATGAGCTTTTGTTCTGAGAACAATTTTTACGCCAAAAAGAGTTGAAGCAAGGATTTGAT			: : : : : : : :		7500
Qy	125	-ProHisLysArgLysSerArgSerLysSerGlnSerLysSerGlnThr			: : : : : : : :		142
Db	7501	AAGCCAGTCGTAAAGACAAATCGTAATCGAAATTCGGCGGTGATGAACAGAAAGTTACT	:	:	: : : : : : : :	:	7560
Qy	143	ArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysLysArgLysAsp	:::	:::	: : : : : : : :	:::	162
Db	7561	AAAGAACACAAACGTCGGAATAA-	:	:	: : : : : : : :	:	7605
Qy	163	ThrLysArgArgGlnGluArgHis-			: : : : : : : :		170
Db	7606	TTTAAATTCGTGAATAATCTACTGACAGAAAACCAAGTTACTGCACAAAATAAAAA			: : : : : : : :		7665
Qy	171	-IleLysGluGluGlnAspLysGlnMetThrSerAlaLysGlnHis			: : : : : : : :		185
Db	7666	GTTTCTGTCAGTAAAAATCTCAAGAGAGAAAAACAGACATAAAAAACAAAA	:	:	: : : : : : : :	:	7719
RESULT 10	AF270167/C	AF270167	3054 bp	DNA	linear	BCT 01-AUG-2000	
LOCUS	Staphylococcus epidermidis strain SRI clone step.1051f01 genomic sequence.						
ACCESSION	AF270167						
VERSION	AF270167.1	GI:9624075					
KEYWORDS							
SOURCE	Staphylococcus epidermidis.						
ORGANISM	Staphylococcus epidermidis.						
REFERENCE	Bacteria; Firmicutes; Bacillales; Staphylococcus.						
AUTHORS	1 (bases 1 to 3054) Kimmerly,W.J., Taylor,J.David., Nelsen,A.J., Godlevski,M.M., Rubino,M.A., Nelson,F.J., Rivers,P.R., Torruella-Miller,I., Listenbee,S., Ashanti,C., Altschuller,G., Mamo,L., Shepherd,N.S., Fuchs,R., Fleming,T., Guan,X., Du,L., Cain,D.H., Miller,G.S. and Furdon,P.J.						
TITLE	Transposon-mediated sequencing of the Staphylococcus epidermidis genome						
JOURNAL	Unpublished						
REFERENCE	2 (bases 1 to 3054)						
AUTHORS	Taylor,J.David., Kimmerly,W.J., Nelsen,A.J., Godlevski,M.M., Rubino,M.A., Nelson,F.J., Rivers,P.R., Torruella-Miller,I., Listenbee,S., Ashanti,C., Altschuller,G., Mamo,L., Shepherd,N.S., Fuchs,R., Fleming,T., Guan,X., Du,L., Cain,D.H., Miller,G.S. and Furdon,P.J.						
TITLE	Direct Submission						
JOURNAL	Submitted (22-MAY-2000) Departments of Genomic Sciences and Bioinformatics, Genetics Directorate, Glaxo Wellcome, Inc., 5 Moore Drive, Research Triangle Park, North Carolina 27709-3398, USA						
FEATURES	source						
	1..3054						
	/organism="Staphylococcus epidermidis"						
	/strain="SRI"						
	/db_xref="taxon:1282"						
	/clone="step.1051f01"						
BASE COUNT	955 a 532 c 440 g 1127 t						
ORIGIN							
Alignment Scores:							
Pred. No.:	9,28e-15	Length:	3054				
Score:	255.50	Matches:	48				
Percent Similarity:	68.22%	Conservative:	25				
Best Local Similarity:	44.86%	Mismatches:	31				
Query Match:	25.32%	Indels:	3				
DB:	1	Gaps:	2				

```
DKLNGVLGDOGLAAWKPLFASAHLDAGAYFTVIIMGALAVIYAKLMLADITIKMPES
VPVAPAKFLAIIPTTAALYIVGLIYIGLKNDSVINLITHYIAEPFOILSONIFS
VLIVTFVSVEFWFGLHGNVLAPVLDGIMPLGLANQALYFOVHSOGLDLIAKAV
DKAHANGDYNLWVRGSDWDAFAGSGGTITLVIAIILFSRKKDYKIVRGLGLAPG
IFNFPVLGDLSPVINAIFBFAVAPLISVIAIYATATALHLVDPVNAVPMWTPPI
MNFAMATGDFDRAIVLTIIINLIITFVIWPFVIAANKLEETELD"
gene
/genes="bgls"
/notes="L179659"
/notes="L179659"
/notes="L179659"
/notes="L179659"
/notes="EVIDENCE BY HOMOLOG Y BIO06.12 ENERGY METABOLISM.
Sugars. HYDROLYSIS OF TERMINAL, NON-REDUCING
BETA-D-GLUCOSE RESIDUES WITH RELEASE OF BETA-D-GLUCOSE"
/codon_start=1
/transl_table=11
/product="beta-glucosidase A (EC 3.2.1.21)"
/protein_id="AAK04273.1"
/db_xref="GI:12723027"
/transl_table=11
/translation="MEHKKLKAPNDPLMGSAAYOVEGAPEDGKKASVWDFVRI
PGKTFKNGDVAVDHYHRYKEDVALMKELGLKSYRESIAWTIRLLPDGRGEVNAQGLK
FYEDLDELIAEIEPIVTIYHMDLQALDELGYGWSREITIAFVNAARVFNAPKMG
KVYVLSNEQIETISQGSWLAHPGPKDKMLFYQVNHIANLTNAVAINKFHELEMD
QIGPSEFATPYAKDSDFNLVLAENAEELYSHFWMDVLYGEYPIAMAYLRENGL
APFEAGDAELKAKPDPLGNGIYQYATNAMPVGGVGVSGFNTSGKGGTKESGI
PGLHKIQNPFIDRTNWDWEIDPEGLRIALRITTSYRIPVMITENGGLGEYDKVEDGK
IHDYRIKYLSEHKAIEKAMTDGAKVIGYHTWSYTDLLSLWNGYKRYGVFVYVDQE
TMGSKLRIPKDSYIYVQHLIETNAEEV"
gene
/genes="dut"
/notes="L181168"
/notes="L181168"
/notes="L181168"
/notes="EVIDENCE BY HOMOLOG Y BIO08.04 Pyrimidine
ribonucleotide biosynthesis. DE NOVO SYNTHESIS OF
THYMIDYLATE. BELONGS TO THE DUTPASE FAMILY"
/codon_start=1
/transl_table=11
/product="deoxyuridine 5'-triphosphate nucleotidhydrolase"
/protein_id="AAK04274.1"
/db_xref="GI:12723028"
/transl_table=11
/translation="MKIRGEVVTYKKNAGINIPKRSTHESAGYDIEAAETVSFAPGE
IKLIPGLKAYMAGEVLYMYDRSNPRKKGLVINSVGVIDKDYNNPNDEGHMFMQ
MRNFTDEEVIEKGERVGVGFVFPFLVADGDNQENXEERTGGFGSTA"
5732..7075
/genes="ybiB"
/notes="L181858"
/notes="L181858"
/notes="L181858"
/genes="ybiB"
/notes="EVIDENCE BY HOMOLOG Y BIO15.01 HYPOTHETICAL.
General. similar to several 5'-nucleotidases"
/codon_start=1
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="AAK04275.1"
/db_xref="GI:12723029"
/transl_table=11
/translation="MKNRLTILNDLHSLHLEFPVIERFEAFKSHSEVESLSFDLGN
VDVHPLTATEQCPNVLMNRLDLTAATIGNNEGILGTHMELSHLYDEANFPVLN
LATNKAQSEPIVTFSGLRIGLIGLTAIPVLAIPQAGWELSDPFVYVDELPLKDCD
FTLLSHLGNVDQIAKDYEDLIDLIAGATHHLEFEGAQVNOVTLAAAGRYGEYVQ
IDITFNENOLVDAQIEATINTLPQRKSDLVETNGWELRGHKLLEGTVVADLPKALP
NVLPDFKSHYLAKIFADYGNTKLKMNSGLVTDLPNHLTLDLPHLTPHSLRIVR
FTFTGBELRVLLEIMDVDFLATOKIQMGFRGKTFGSLIFHGIEPRGDFYIKNSE
NQSLEKISDEESQIVLPDQYLFANWYFPLKLKCKSEILFPFLREIVAEHFNK"
7090..7845
/genes="ybiC"
/notes="L183216"
/notes="L183216"
/notes="L183216"
/genes="ybiC"
/notes="EVIDENCE BY HOMOLOG Y BIO15.01 HYPOTHETICAL.
General. 53% identical to Yutd protein of B.subtilis"
/codon_start=1
/transl_table=11
```

```
/product="HYPOTHETICAL PROTEIN"
/protein_id="AAK04276.1"
/db_xref="GI:12723030"
/transl_table=11
/translation="MAKVIDESKLNYNKYPGEHVMAGVEVVGQRTFHTLVHNYREAF
DAKLQORSDVLDKYDIIVGDMGFEQLRKGFSTSRKMLADNKIDHLDEQVYNEFC
NYGCAVEYIRIRTKDEAFYKELFGKGFDPKRRKRNRRNRNWARDQKVTKEED
KTSNSEARKKDFKIREKSTDRPKPKVTDKNKYSVSKSQERKTDNKKQNPADSKKK
FTMKRNPRKKVEQTETTKQENQGFVLRNKKD"
8039..9136
/genes="ybiD"
/notes="L184159"
/notes="L184159"
/notes="L184159"
/genes="ybiD"
/notes="EVIDENCE BY HOMOLOG Y BIO15.01 HYPOTHETICAL.
General. conserved hypothetical protein"
/codon_start=1
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="AAK04277.1"
/db_xref="GI:12723031"
/transl_table=11
/translation="WTENIITETRPSIYGLTRDLIEWAIENGKEKKERATQVWDWLYR
KRVQSEEMSNSAAFTDKLNEHFIPLNEQVVSQSDAGTVKYLFLPQMIMETVL
MRQSYGLSVCTVTVQCNMGCTFCASGLKKEKRDVTAGETVSIQMLVKYKDEGLDE
RVSHVVMGIEPFDNYEHLNMFRLVINDDNGLAIGHARITVTCGCFMPAKIKEFAHE
NLQINLAISLHAPNNEELRTSLMRTNAPLEKLEAFIDYVTTNRRVTVYIIMLGE
NDSPETAQOLADLIKPRNKLSVNLIPYNPVAEHIKYERSTKNTAKFYDVLKKNGIN
CVVRQEHQTDIDAACGQLRSKQIKKNKAKLA"
9382..9993
/genes="ybiE"
/notes="L185517"
/notes="L185517"
/notes="L185517"
/genes="ybiE"
/notes="EVIDENCE BY HOMOLOG Y BIO06.01 ENERGY METABOLISM.
Aerobic. similar to different oxidoreductases"
/codon_start=1
/transl_table=11
/product="oxidoreductase"
/protein_id="AAK04278.1"
/db_xref="GI:12723032"
/transl_table=11
/translation="MDYVKLNKRHRKKNFDGKKIPTVDVKQIISASLAPSAHNTQS
WHFVIVSEKREKREALLSEVDPGNHQIKAGAVILFSDTDLVERSDRARGAGELD
DQQLFRNRYRQMYEGFELHESYLSINIGLVTNLVLAIKNYGESNIIMGFNRT
QKVNLEILEYKRYRPELIIPLGNSDKGKPSYRLPQSQIMEIR"
BASE COUNT 3369 a 1669 c 1992 g 3068 t
ORIGIN

Alignment Scores:
Pred. No.: 2.58e-21 Length: 10098
Score: 331.50 Matches: 75
Percent Similarity: 51.83% Conservative: 38
Best Local Similarity: 34.40% Mismatches: 64
Query Match: 32.85% Indels: 41
DB: 1 Gaps: 5

US-09-868-352-23 (1-192) x AE006255 (1-10098)
Qy 1 MetArgLysGluValThrProGluMetLeuAsnTyrAsnLysTyrProGlyProGlnPhe 20
Db 7090 ATGCTAAAGTAATAGACGAAGTAAGTTAAATTAATATAATATCTCGCGAACATGTT 7149
Qy 21 IleHisPheGluAsnIleValLysSerAspAspIleGluPheGlnLeuValIleAsnGlu 40
Db 7150 ATGCACATAGGGGAAGTGTTCAGTTGGACACAGCACTTTCATATTTTCATAATAT 7209
Qy 41 LysSerAlaPheAspValThrValPheGlyGlnArgPheSerGluLeuLeuLysTyr 60
Db 7210 CGTGAGCGCTTTTGATGCGGAAAAATTAGAACACAGCACTTTTCTGATGTTTAGATAAGTAT 7269
Qy 61 AspPheIleValGlyAspTTPGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAsp 80
Db 7270 GATTATATTCGGAGATTCGGGCTTTTGACCAATTCGCTTTTGAAGGCTTTTTC----- 7323
Qy 81 AlaSerThrIleArgLys-----AsnSerArgIleSerArgLeuGluAspTyr 96
```


MEDLINE PUBMED REFERENCE AUTHORS	21357209 11463916 2 (bases 1 to 10029) Tettelin,H., Nelson,K.E., Paulsen,I.T., Eisen,J.A., Read,T.D., Peterson,S., Heidelberg,J., Deboy,R.T., Haft,D.H., Dodson,R.J., Durkin,A.S., Gwinn,M., Kolonay,J.F., Nelson,W.C., Peterson,J.D., Umayam,L.A., White,O., Lewis,M.R., Radune,D., Holtzapple,E., Khouri,H., Wolf,A.M., Utterback,T.R., Hansen,C.L., McDonald,L.A., Feldblyum,T.V., Anquioli,S., Gesuwan,P., Hickey,E.K., Holt,I.E., Loftus,B.J., Ujwal,M.L., Yang,F., Smith,H.O., Venter,J.C., Dougherty,B.A., Morrison,D.A., Hollingshead,S.K. and Fraser,C.M. Direct Submission Submitted (29-JUN-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA	gene CDS	/translation="MAIILPELPAYDALEPYIDAETMHLHDKHQTQVYNNANAALAE KHEIGEDLEALLADVESYAPADIROALINNGGHLNHALFWELMTPEKTAPSAELAAA IDATGSGFEFOAAATAATBFGSCHWALVYNKCKLEVSTANQDTPISEKCKPIL GLDVEHAYVYKRVRPDIKAFPSVINWKNVDELYAAK" 3418..3930 /gene="SP0767" 3418..3930 /gene="SP0767" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAK74905.1" /db_xref="GI:14972242" /translation="MRKEIAPELYNKPEPGFEHLHGDKVETEGIAFSLVENIKDAF DVTTFNORFSEVLTKFDYIVGWSNEQLRGRFYKDRTEEKLEKISRLODYLLCYCS YGCAYFVLENEAPKRASFDDKMKRKEETPSRKKKPTQTQKRSNADKKNRNRQRDKK GCRHFVIRQK" 3951..5036 /gene="SP0768" 3951..5036 /gene="SP0768" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAK74906.1" /db_xref="GI:14972243"
	1..10029 /organism="Streptococcus pneumoniae TIGR4" /strain="TIGR4" /db_xref="taxon:170187" 216..344 /gene="SP0763" 216..344 /gene="SP0763" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAK74901.1" /db_xref="GI:14972238" 478..1413 /gene="SP0764" 478..1413 /gene="SP0764" /note="identified by match to PFAM protein family HMM PF01180" /codon_start=1 /transl_table=11 /product="dihydroorotate dehydrogenase A" /protein_id="AAK74902.1" /db_xref="GI:14972239" /translation="MVSTKTQTAGFEFONCLMNAAGVACMTIEELEEVKNSAAGTFTV KTATIDFRGNPEPRYQDVPGLSINSMGLPNNGLDYLDYLDLOEKESNRFTFLSLV GMSPEPTHILAKVQESDPRGLTNLSCPVPKQPIAYDEFTDRIAEVFAVFTK PLGIKLPPYDVHFDQAAIFNKVPLKPVNCVNSIGNGLYIEDSVVIRPKNGGGI GGEYIKPTALANHAFYQRLNPQIQIIGTGGVLTGRDAFEHLCLGASMVQVGTTLHKE GVSAFDRTNEUKALMVEAGYESLEDFRGLRTID" 1447..2484 /gene="SP0765" 1447..2484 /gene="SP0765" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAK74903.1" /db_xref="GI:14972240" /translation="MLAIEESOKLTLSNLPISLFTGTDOGFQFVNMKSOMLQICGYDS ADLNFRAYFMKEVYKDVLELVSLPFFADEKIVLDVPMFTTAKKRLTDDDELKSF EYTLNPSPTTKLIIFAEGKLDKSRRLKLRDAKAFDAVEKOEULRQYFQKWSQF QSLQTHNSFDLTKSGFQSEIQKNLLFLQSYKANSYIEEDIVNAIPKTDQNIIF DITQFTLTKMQDQARDLVRLDTLQGEDEIKLIAMVLGQPTFTQVKILAESGOTESQI ASSLGSYLRGNPNPQIKFALRDSRGLSLFLKQALISYLIETDYQIKTGLYKGFIFE KALLQIASOVN" 2657..3262 /gene="SP0766" 2657..3262 /gene="SP0766" /codon_start=1 /transl_table=11 /product="superoxide dismutase, manganese-dependent" /protein_id="AAK74904.1" /db_xref="GI:14972241"		
FEATURES source	/translation="MAIILPELPAYDALEPYIDAETMHLHDKHQTQVYNNANAALAE KHEIGEDLEALLADVESYAPADIROALINNGGHLNHALFWELMTPEKTAPSAELAAA IDATGSGFEFOAAATAATBFGSCHWALVYNKCKLEVSTANQDTPISEKCKPIL GLDVEHAYVYKRVRPDIKAFPSVINWKNVDELYAAK" 3418..3930 /gene="SP0767" 3418..3930 /gene="SP0767" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAK74905.1" /db_xref="GI:14972242" /translation="MRKEIAPELYNKPEPGFEHLHGDKVETEGIAFSLVENIKDAF DVTTFNORFSEVLTKFDYIVGWSNEQLRGRFYKDRTEEKLEKISRLODYLLCYCS YGCAYFVLENEAPKRASFDDKMKRKEETPSRKKKPTQTQKRSNADKKNRNRQRDKK GCRHFVIRQK" 3951..5036 /gene="SP0768" 3951..5036 /gene="SP0768" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAK74906.1" /db_xref="GI:14972243"	gene CDS	/translation="MKPSIHSLSHVHOTMOEWLVLEQGEKKFRAQOIWMLYKRVQSPFEE MTNLKDIKALNDQFVNPFLKORIVQESADGTVYVLFELPDGMLETIYLMRQHYGLS VCVTTQVGCNIGCTFCASGLIKKQDLNNGEIVAOIMLVQKIFYEDGQDERISHIVVM GIGEPFDYNNVNFRTINDDKMAIGARHITVSTGLAHKIRDFADREGVGNLAVS LHAPNELRSSIMKINRAFFIEKLFAAIEYIETNRRVTPEYIMLNEYNDGVEQALE LTELAKNLKLSIVNLIPYNVSEHDOISRSPECVLAIFYTLKKKGVCVVRQEHGT DIDACGQLRSNTMKRDRQKAAVAVNP" 5039..5518 /gene="SP0769" 5039..5518 /gene="SP0769" /note="similar to GP:7110140; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="VanZP-related protein" /protein_id="AAK74907.1" /db_xref="GI:14972244" /translation="MKTTYNHILVMGVIPYSICVCFCTPQEQSTQVGVGTGPIQHLG RLVFLTFPNSLWMLGVSVDIGOLCMIFLQNLINLVFPFLIFQLLYLPNLRKTKV LLFSLVSLGIECTQLILDFFDFNRFVEIDDLWTNTLGGYLAWLLYKRLHKNKVRN" 5520..7061 /gene="SP0770" 5520..7061 /gene="SP0770" /note="ABC transporter, ATP-binding protein" /product="ABC transporter, ATP-binding protein" /protein_id="AAK74908.1" /db_xref="GI:14972245" /translation="MSILEVKNLSHGFGDRAIPEDVSFLLKGEHIGLVGANGEGKST FNSIVTGMPLDGEKVSQYTAGYLDQHSYLAERQSVRDLVIRAFDELFAEARIN DLYMKMAGDADALMEVEGELQDLRESDFYTLDAKIDEVARALGVMDFGMDTDTVT SLSGQRTKVLKALLLEKPDILLDEPTNYLDAEHIQDWLRYLQNYENAFVLISHDI PFLNDVINIVHVENQOLTRYSDIYQFOEVYAMKKSOLEAAVROOKEIADLKDFVA RNKARVATRNAMMSRQKLDKMDIIELOSEKPKPSDFKPARTPGFIFQAKNLQIGY DRPLTKPLNLTFRNQKVAIICANGIKGTTLTKSLGIIISPIAGEVERGDYLELGYFE QVEGNGRQTPLEAVNAPALNQAEVRAALRCGLTTKTHIESQIQVLSGGGQAKVRF CLLMNRNNVLVDEPTNHLVDVDAKDELKRALKEYRGSLMVMVCFEPDFVEGDIQIWD FNLLT" complement(7117..7920) /gene="SP0771" complement(7117..7920) /gene="SP0771" /codon_start=1

ORIGIN

Alignment Scores:
Pred. No.: 9.29e-32 Length: 5558
Score: 438.00 Matches: 88
Percent Similarity: 68.89% Conservative: 36
Best Local Similarity: 48.89% Mismatches: 50
Query Match: 43.41% Indels: 6
Db: Gaps: 4

US-09-868-352-23 (1-192) x BD003783 (1-5558)

Qy 1 MetArgLysGluValThrProGluMetLeuAsnTyraSnLysTyrPrroGlyProGlnPhe 20
|||||
Db 4895 ATGCCAAAGAAATTCACCTGAATTATACAATATACAAAGTTCCGTGGCGGAGTTC 4826

Qy 21 IleHisPheGluAsnIleValLysSerAspIleGluPheGlnLeuValIleasnGlu 40
|||
Db 4825 CATTACACGGGGCAGAAGTCGAAACGGAAGGATAGCTTTTCCCTGGTTGAAAATATC 4766

Qy 41 LysSerAlaPheAspValThrValPheGlyGlnArgPheSerGluIleLeuLysTyr 60
|||||
Db 4765 AAGGATGCCTTGATGTGACGACTTTTAATCAGCGTTTTTTCAGAAGTATTACCAAGTTT 4706

Qy 61 AspPheIleValGlyAspTrpGlyAsnGluInLeuArgLeuArgGlyPheTyrLysAsp 80
|||||
Db 4705 GATTATATCGTGGGGACTGGAGCAACGACAGCTTCGCTACGAGGTTTTTTACAAGGAT 4646

Qy 81 AlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyr 100
|||||
Db 4645 GACCGAACAGAGAAAACTTGAAAATAATCATGCTGTTTACAAGACTACCTTTTAGAGTAT 4586

Qy 101 CysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspIleLysPhe 120
|||||
Db 4585 TGTAGTTATGTTGTGCCTATTTTGTCTAGAAAATGAAGCCCCTAAGCGAGCATCATTT 4526

Qy 121 AspaspGluArgProHisLysArgGlySerArgSerLysSerGlnSerserLysSer 140
|||
Db 4525 GAC-----AAGAAATCGTAAGACGGAAGAAACACCCTCTAGAAAAGGAAGAA 4472

Qy 141 GlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysLysArg 160
|||||
Db 4471 CCGACTCAAACCTAAACGA--NAGTCGAATGCAGATAGAAA--NATGACGTCGTCTAG 4418

Qy 161 LysAspThrLysArgArgGlnGluArgHisIleLysGluGlu-----GlnAspLysGlu 178
|||||
Db 4417 AAAGACGACAAAGGACAAACGTCATTTTGTCTAGAGTTCGTCAGAGTAGTAGAGAATAAGGAG 4358

RESULT 8
AE007384
LOCUS AE007384 . 10029 bp DNA linear BCT 31-AUG-2001
DEFINITION Streptococcus pneumoniae TIGR4 section 67 of 194 of the complete genome.
ACCESSION AE007384 AE005672
VERSION AE007384.1 GI:14972237
KEYWORDS
SOURCE Streptococcus pneumoniae TIGR4.
ORGANISM Streptococcus pneumoniae TIGR4
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
1 (bases 1 to 10029)
AUTHORS Tettelin,H., Nelson,K.E., Paulsen,I.T., Eisen,J.A., Read,T.D., Peterson,S., Heidelberg,J., DeBoy,R.T., Haft,D.H., Dodson,R.J., Durkin,A.S., Gwinn,M., Kolonay,J.F., Nelson,W.C., Peterson,R.J., Hayam,L.A., White,O., Salzberg,S.L., Lewis,M.R., Radune,D., Holtzapple,E., Khouri,H., Wolf,A.M., Utterback,T.R., Hansen,C.L., McDonald,L.A., Feldblum,T.V., Anquili,S., Dickinson,T., Hickey,E.K., Holt,I.E., Loftus,B.J., Yang,F., Smith,H.O., Venter,J.C., Dougherty,B.A., Morrison,D.A., Hollingshead,S.K. and Fraser,C.M.
Complete genome sequence of a virulent isolate of Streptococcus pneumoniae
Science 293 (5529), 498-506 (2001)
JOURNAL

QVEVGGNRQTPLEAVNNAFFALNQAEVRAALARCGLTTKKHIESQIOVLGSGEQAKVRF
CLLMRENNNVLVLDEPTNHLNDVDAKDELKRALKEYRGSILMVCHEPDPFYECWIDQIWD
FNNLT"
gene complement(7182. .7985)
/gene="ppia"
/note="spr0679"
CDS complement(7182. .7985)
/gene="ppia"
/note="putative"
/codon_start=1
/transl_table=11
/label_spr0679
/product="Peptidyl-prolyl cis-trans isomerase"
/protein_id="AAK99483.1"
/db_xref="GI:15458267"
/translation="MKLATLIIIIITVALACSSVQSRSLRGDDYVDSLLAERSSKVA
AQSAKELNDALTENAFNPQLSKEVDEAEVLIHTSOGDIRIKLFPKPLAVENFL
THAKGYNGITTHRVLDGFMVGTGQGSIMHDKDKTKDGTGFKNEITPY
LYNIRGALAMANTQPNNGSQFFINQNSDTSSKLPSTKYPQRIIEAYKEGKNPSLD
GKHPVQGVIGGMDVVDKIAKAEDKDKPTTATTIDSEVVKDYDFKS"
gene complement(8069. .8290)
/gene="spr0680"
CDS complement(8069. .8290)
/gene="spr0680"
/codon_start=1
/transl_table=11
/label_spr0680
/product="Hypothetical protein"
/protein_id="AAK99484.1"
/db_xref="GI:15458268"
/translation="MSFYGLFYNGIAITPNTYLSAFVNFVETALPLNLFIVEPIARI
LSSFKPFTGEEVEEVEDFDDDEIPTII"
gene complement(8178. .8408)
/gene="spr0681"
CDS complement(8178. .8408)
/gene="spr0681"
/codon_start=1

Alignment Scores:
Pred. No.: 8,27e-33 Length: 10029
Score: 452.50 Matches: 94
Percent Similarity: 66.49% Conservative: 33
Best Local Similarity: 49.21% Mismatches: 48
Query Match: 44.85% Indels: 16
DB: 1 Gaps: 4

US-09-868-352-23 (1-192) x AE008445 (1-10029)

Qy 1 MetArgLysGluValThrProGluMetLeuAsnTyrAsnLysTyrProGlyProGlnPhe 20
Db 3477 ATGCGAAAAGAAATGACCTGAATTATACAACTATACAAAGTTCCGTGTCGGAGTTC 3536
Qy 21 IleHisPheGluAsnIleValLysSerAspIleGluPheGlnLeuValIleAsnGlu 40
Db 3537 CATTTACACGGGACAGAGTCGAAACGGAAGGATAGTATTTTCTGTTGTAATAATC 3596
Qy 41 LysSerAlaPheAspValThrValPheGlyGlnArgPheSerGluIleLeuLysTyr 60
Db 3597 AAGGATGCTTTGATGTGACGACTTTTAATCAGCGTTTTTTCAGAAAGTATTACCAAGTTT 3656
Qy 61 AspPheIleValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAsp 80
Db 3657 GATTATATCGTGGGACTGGAGCAACGACAGCTTCGCCCTACGAGGTTTTTACAAGAT 3716
Qy 81 AlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyr 100
Db 3717 GACCGAACAGAGAAAACATGAAAAATCAGTCTTTTACAAGACTACCTTTTACAGAT 3776
Qy 101 CysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspIleLysPhe 120
Db 3777 TGTAGTTATGGTGTGCGCTATTTGTCTTAGAAATGAAGCCCTTAAGCGGAGCATCATTT 3836
Qy 121 AspAspGluArgProHisLysArgLysSerArgSerLysSerGlnSerSerLysSer 140

Db 3837 GAC-----AAGAAATCGCTAAGCAGGAAGAAACACCTTCTAGAAAAGGAAGAA 3890
Qy 141 GlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysLysArg 160
Db 3891 CCGACTCAAACTAAACGA---AAGTCGAATCGAGT----- 3923
Qy 161 LysAspThrLysArgArgGlnGluArgHisIleLysGluGlnAspLysGluMetThr 180
Db 3924 AAGAAATAGACGCTGCAGAAAGACCAAGCATTTCTCAGAAAGAGGACAAAGAA----- 3977
Qy 181 SerAlaLysGlnHis-LeuLeuPheValArg 190
Db 3978 -----CAACGTCATTTTGTTCATTCGTCAGA 4002
RESULT 6
SPNEU1904 147006 bp DNA linear HTG 11-JUL-2001
LOCUS Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***,
DEFINITION in ordered pieces.
ACCESSION AL449926
VERSION AL449926.1. GI:11545151
KEYWORDS HTG; HTGS_PHASE2.
SOURCE Streptococcus pneumoniae.
ORGANISM Streptococcus pneumoniae.
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 147006)
AUTHORS Dopazo,J., Mendoza,A., Herrero,J., Caldara,F., Polissi,A.,
Friedli,L., Guerrier,M., Grand-Schenk,E., Gandin,C., de
Francesco,M., Polissi,A., Buell,G., Feger,G., Garcia,E., Peitsch,M.
and Garcia-Bustos,J.F.
TITLE Annotated draft genomic sequence from a Streptococcus pneumoniae
JOURNAL type 19F clinical isolate
MEDLINE Microb. Drug Resist. 7 (2), 99-125 (2001)
PUBMED 21335329
REFERENCE 11442348
AUTHORS 2 (bases 1 to 147006)
Dopazo,J., Mendoza,A., Herrero,J., Caldara,F., Polissi,A.,
Humbert,Y., Friedli,L., Guerrier,M., Grand-Schenk,E., Gandin,C., de
Francesco,M., Buell,G., Feger,G., Garcia,E., Peitsch,M. and
Garcia-Bustos,J.F.
TITLE Direct Submission
JOURNAL Submitted (31-OCT-2000) Research Department, Glaxo Wellcome, S.A.,
Severo Ochoa 2, 28760 Tres Cantos, SPAIN
COMMENT * NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source 1. 147006
/organism="Streptococcus pneumoniae"
/serotype="19F"
/db_xref="taxon:1313"
/clone="G54"
BASE COUNT 44762 a 26837 c 32084 g 43281 t 42 others
ORIGIN
Alignment Scores:
Pred. No.: 2,52e-31 Length: 147006
Score: 452.50 Matches: 94
Percent Similarity: 66.49% Conservative: 33
Best Local Similarity: 49.21% Mismatches: 48
Query Match: 44.85% Indels: 16
DB: 2 Gaps: 4
US-09-868-352-23 (1-192) x SPNEU1904 (1-147006)
Qy 1 MetArgLysGluValThrProGluMetLeuAsnTyrAsnLysTyrProGlyProGlnPhe 20
Db 106422 ATGCGAAAAGAAATGACCTGAATTATACAACTATACAAAGTTCCGTGTCGGAGTTC 106481
Qy 21 IleHisPheGluAsnIleValLysSerAspIleGluPheGlnLeuValIleAsnGlu 40


```

/ gene="SpyM3_1152"
/ complement(5346..5816)
/ gene="SpyM3_1152"
/ note="best blast hit: gb|AAK34295.1| (AE006583) putative
repressor protein [Streptococcus pyogenes M1 GAS], and
gb|AAL98085.1| (AE010066) putative repressor protein
[Streptococcus pyogenes MGAS8232]"
/ codon_start=1
/ trans_table=11
/ product="putative repressor protein"
/ protein_id="AAM79759.1"
/ db_xref="GI:21904891"
/ translation="MKKSERLEIKKMWLTHPIETQHDLLRLLAEGHLELTQATISRD
PNETGIKVPISGRIYIGLSQDSKRIYGVGPRSIKSTILAVSDKTKGLBQHLVKKV
MNGSKLTKRYLLADFSKAIFSLIADDDSLLLIAKSPSEADMIRQEIILLWQGIT"
/ complement(5803..6630)
/ gene="hlyA"
/ note="hlyA"
/ gene="SpyM3_1153"
/ complement(5803..6630)
/ gene="hlyA"
/ note="best blast hit: gb|AAL98086.1| (AE010066) putative
hemolysin [Streptococcus pyogenes MGAS8232]"
/ codon_start=1
/ trans_table=11
/ product="putative hemolysin"
/ protein_id="AAM79760.1"
/ db_xref="GI:21904892"
/ translation="MPKERVDLVAYKQGLPETREQAKRGVMAGLVVSVINGQRYDKPG
DKINDGTELKLGKLYKVSRGGLKLEGLHVFQVSVANQIGIDIGASTGFTDVMLO
DGAKLVADVDTGNOLVWKURQDPRVSRMEQYFNRYAQEDFSEGVFASIDVSFIS
LSLPLPALHNSDQGVIALIKPFQFAGREQIGKGIKVKDQIHERKVIQKVMDFASG
YGVTFVKDFPSIOGGHGNIEFLAHLVKSQTPETLAPHLIQEVVAKHKEFEKHEK"
/ complement(6623..7495)
/ gene="fps"
/ note="SpyM3_1154"
/ complement(6623..7495)
/ gene="fps"
/ note="farnesyl diphosphate synthase; best blast hit:
gb|AAL98087.1| (AE010066) putative geranyltranstransferase
[Streptococcus pyogenes MGAS8232]"
/ codon_start=1
/ trans_table=11
/ product="putative geranyltranstransferase"
/ protein_id="AAM79761.1"
/ db_xref="GI:21904893"
/ translation="MDKLARIDEAIRRYKTTNSGVSEELIDAILLYSDSGGKRIRPL
ILLEMIGFVSLQNAHFDAALAEMLHTGSLIHDDLPMNDYRRGRLLNNHKQFGE
ATAILAGDSFLDPFGLIQAQELNSEYKVALIQELSLASGTFGVMGQMLDMKQENQA
LSLPQLSLIHLNKTGKLLTFPFAAALVTQAITVROOLEQAGMLIGHAFQIRDDILD
VTASFDELGTPKPKDLFAERATYPSLLGLEASYQLLTESLDQALTIFOTLENDVGFKP
QITPKLTGRLNA"
/ complement(7495..7710)
/ gene="xseB"
/ note="SpyM3_1155"
/ complement(7495..7710)
/ gene="xseB"
/ note="best blast hit: gb|AAK34298.1| (AE006583)
exodeoxyribonuclease VII (small subunit) [Streptococcus
pyogenes M1 GAS], and gb|AAL98088.1| (AE010066)
exodeoxyribonuclease VII (small subunit) [Streptococcus
pyogenes MGAS8232]"
/ codon_start=1
/ trans_table=11
/ product="putative exodeoxyribonuclease VII small subunit"
/ protein_id="AAM79762.1"
/ db_xref="GI:21904894"
/ translation="MSKTRTFEENLQDLETIVNKLKENGDPVLEEAISEFQKGMLLSRE
LQKTLQAAEKTLLVKMQADGTEVDMDD"
/ complement(7688..9028)
/ gene="xseA"
/ note="SpyM3_1156"
/ complement(7688..9028)

```

```

Alignment Scores:
Pred. No.: 6,91e-42 Length: 50463
Score: 557.50 Matches: 107
Percent Similarity: 71.98% Conservative: 42
Best Local Similarity: 51.69% Mismatches: 41
Query Match: 55.25% Indels: 17
DB: 1 Gaps: 4

US-09-868-352-23 (1-192) x AE014159 (1-50463)

Qy 1 MetArgLysGluValThrProGluMetLeuAsnTyrAsnLysTyrProGlyProGlnPhe 20
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 39611 ATGAAAAAGAAATTTCCGCAAGAAATGTACAACTATAATAATTTCCAGGTCCTCAATTT 39552
Qy 21 IleHisPheGluAsnIleValLysSerAspIleGluPheGlnLeuValIleAsnGlu 40
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 39551 ATTCAATTTGAAGCAAGTTAAGGCTGAAGCATTTGTTACTCTTAGAAGATGTT 39492
Qy 41 LysSerAlaPheAspValThrValPheGlyGlnArgPheSerGluIleLeuLeuLysTyr 60
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 39491 AAGAACGCTTTTGATACGACTAGCTTTGTCACACGTTATACAGAAAGTTCTCCTAAGATAT 39432
Qy 61 AspPheIleValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAsp 80
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 39431 GACTATATTGTTGGCGATTGGGGAATCAACAGCTTCGCTCTTAAAGGCTTTTATAAGGAT 39372
Qy 81 AlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyr 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 39371 AGTGATGATATACAGAAACCAATCGCATCTCACGTTTAGAAGATATATTAAAGAAATT 39312
Qy 101 CysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspIleLysPhe 120
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 39311 TGCATTTTGGTGTGCTTATTGTCCTAGAAAATCTTCATCCACAGATATTAAATTT 39252
Qy 121 AspAspGluArgProHisLysArgArgLysSer---ArgSerLysSerGln----- 136
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 39251 GAAGAGGAGCGCCACCAACCAAGACGAGAAAGAAATCACCTAATCAAAATCAATCGTCGCAAG 39192
Qy 137 -----SerSerLysSerGlnThrArgAsnAsn 145
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 39191 CCAACTATTCAATCAGCCGCCCAACACCTAGAGATTAATCGAAG---CGGTATCA 39135
Qy 146 ArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysLysArgLysAspThrLysArg 165
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 39134 AAAGAAAGCAACCTGAAACCAACCAACCGTTTACCAGTCAAAACGTCGAAGTAATACTAAG 39075
Qy 166 ArgGlnGluArgHisIleLysGluGluGlnAspLysGluMetThrSerAlaLysGlnHis 185
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 39074 CATAAAGAAAAGTCA---AAACGTAATCAGACTAGTCAACTTAATACCAAAATTTAGTCAT 39018
Qy 186 LeuLeuPheValArgLysAsn 192
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 39017 TTTATCATCAGAAAGAAAGAT 38997

RESULT 5
AE008445 10029 bp DNA linear BCT 13-SEP-2001
LOCUS Streptococcus pneumoniae R6 section 61 of 184 of the complete
DEFINITION genome.
ACCESSION AE008445 AE007317
VERSION AE008445.1 GI:15458259
KEYWORDS
SOURCE Streptococcus pneumoniae R6.
ORGANISM Streptococcus pneumoniae R6.
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 10029)
AUTHORS Hoskins,J.A., Alborn,W. Jr., Arnold,J., Blaszcak,L., Burgett,S.,
DeHoff,B.S., Estrem,S., Fritz,L., Fu,D.-J., Fuller,W., Geringer,C.,
Gilmour,R., Glass,J.S., Khoja,H., Kraft,A., Lagace,R.,
LeBlanc,D.J., Lee,L.N., Lefkowitz,E.J., Lu,J., Matsushima,P.,
McAhren,S., McHenney,M., McLeaster,K., Mundy,C., Nicas,T.I.,
Norris,F.H., O'Gara,M., Peery,R., Robertson,G.T., Rockey,P.,

```

QY 137 -----SerSerLysSerGlnThrArgAsnAsn 145
 Db 7071 CCAAACTATTCAAATCAGCAGCCGCAACACCTAAGAGTAATCTGAAG---CGGTATCA 7015
 QY 146 ArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysLysArgLysAspThrLysArg 165
 : : : : :
 Db 7014 AAGAAAGCAACTGAAACAGCGTTTACCAGTCAAAACGTCGAAGTAATACTAAG 6955
 : : : : :
 QY 166 ArgGlnGluArgHisLeIleLysGluGlnAspLysGluMetThrSerAlaLysGlnHis 185
 : : : : :
 Db 6954 CATAGAAAGTCA---AAACGTATCAGACTAGTCAACTTAATACCAAAATAGTCAT 6898
 : : : : :
 QY 186 LeuLeuPheValArgLysAsn 192
 : : : : :
 Db 6897 TTTATCATCAGAAAGAAAGAT 6877

RESULT 4
 AE014159/c

LOCUS
 DEFINITION Streptococcus pyogenes MGAS315, section 24 of 37 of the complete genome.
 ACCESSION AE014159
 VERSION AE014159.1
 KEYWORDS GI:21904885
 SOURCE
 ORGANISM Streptococcus pyogenes MGAS315.
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 Streptococcus.
 1 (bases 1 to 50463)
 Beres,S.B., Sylva,G.L., Barbican,K.D., Lei,B., Hoff,J.S.,
 Mammarella,N.D., Liu,M.-Y., Smoot,J.C., Porcella,S.F.,
 Parkins,L.D., McCormick,J.K., Leung,D.Y.M., Schlievert,P.M. and
 Musser,J.M.

REFERENCE
 AUTHORS
 TITLE Genome sequence of a serotype M3 strain of group A Streptococcus:
 Phage-encoded toxins, the high-virulence phenotype, and clone
 emergence

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (15), 10078-10083 (2002)
 REFERENCE 2 (bases 1 to 50463)
 AUTHORS Beres,S.B., Sylva,G.L., Barbican,K.D., Lei,B., Hoff,J.S.,
 Mammarella,N.D., Liu,M.-Y., Smoot,J.C., Porcella,S.F.,
 Parkins,L.D., McCormick,J.K., Leung,D.Y.M., Schlievert,P.M. and
 Musser,J.M.

TITLE Direct Submission
 JOURNAL Submitted (14-JUN-2002) Laboratory of Human-Bacterial Pathogenesis,
 Rocky Mountain Laboratories, NIAID, NIH, 903 South Fourth St.,
 Hamilton, MT 59840, USA

FEATURES
 SOURCE
 1. 50463
 /organism="Streptococcus pyogenes MGAS315"
 /strain="MGAS315"
 /serotype="M3"
 /db_xref="taxon:198466"
 /notes="group: A"
 complement(80..667)
 /gene="SpyM3_1147"
 complement(80..667)
 /gene="SpyM3_1147"
 /notes="best non-GAS blastp hit: gb|AAK04599.1|AE006285.7
 (AE006285) hypothetical protein [Lactococcus lactis subsp.
 lactis], and gb|AAK08221.1|AF320916_2 (AF320916) Yfaa
 [Lactococcus lactis]"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAM79754.1"
 /db_xref="GI:21904886"
 /translation="MKKSNLNNKWSFICLLAFNFAFLMVITASRLIQVREPESELIA
 KPVKNKIGTFTTREQNETVASYLKDYQTKMKSYKFTYSSSILFEGTQVLGYE
 IVYNLQIDQNDKRVQVLSFSVGTLPLEPKVDVLQYKLKSYKLPSEFKVMPNQSA
 complement(645..1487)
 /gene="SpyM3_1148"
 complement(645..1487)

/gene="SpyM3_1148"
 /note="best blastp hit: emb|CAA72096.1| (v11213)
 hypothetical protein [Streptococcus thermophilus]"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAM79755.1"
 /db_xref="GI:21904887"
 /translation="MNNRHLFSGIPFFVISLCIAFLLLNIIIPKNSRLKKSDFLKKE
 QVAIQVAIGDSLTGEGVGLTHQGGFVPLLTINDLSEYFRKANNHONGYVSGDTSQIL
 DRMIKQQLSLKRAIDMTLVGGNDVMVIRKNUADLQVSRFRPAQRIKQRI
 IELARKNDKDFIFILGTYNPFYLPNPELTDMOKVIDDNTTKTEVGYDRYFVPI
 NDLLYKINGIQEGIVHSSGDQTTIVNDALFTGDHFHPNNTGTQIMSNVMEKIKKEK
 KIKP"
 complement(1480..2319)
 /gene="SpyM3_1149"
 complement(1480..2319)
 /gene="SpyM3_1149"
 /note="best non-GAS blastp hit: gb|AAK75223.1| (AE007412)
 degv family protein [Streptococcus pneumoniae TIGR4]"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAM79756.1"
 /db_xref="GI:21904888"
 /translation="MGTIKIVTDSITIEPELIKALDITVPLVMSIDSKLSYSDNDLK
 EEGHFLSLMKASKSLPKTSQPPVGLFAETIENLVKGVTDIVAIHLSPALSGTIEASR
 QGAETAEPTVTLDSGFTQAMKFOVVEAKWAKAGASINELIILAAQAIKSKTELVI
 VSTLENLVKGGRIQVGTGVLSSLLNVKVMALKNDKELTLVKGKQTKTKWDSVLA
 KNSHRPIAIEIAISYAGEASLALTKERIAAYNHISISVLETSGSIITHTHGEGAFVMV
 RYE"
 complement(2547..3491)
 /gene="SpyM3_1150"
 complement(2547..3491)
 /gene="SpyM3_1150"
 /note="best blastp hit: gb|AAL98083.1| (AE010066)
 hypothetical protein [Streptococcus pyogenes MGAS8232]"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAM79757.1"
 /db_xref="GI:21904889"
 /translation="MTSKRACLSSIIVLASLTCGNQTVTSANHLSATGCDKFDGCGSTLVE
 KIVAPKDELEMLAWSSQTTDDADRYEDFLDDDSFNDDSFNDDSFISQNETDKMFEN
 LTRDLRLDEDEDTIEPQNVIMPDSDELDTDAVETRLTVSSAPHLAEALPKPHL
 RSLSDTLRSGEIRGHLDKDLASVTATKALTMQKFDLTHTVYSIGESFSEVIAA
 HYEDRKAESAFAKSKRFLPIATPDVVIEELRLVSSIGSKEDVSVYPYNRKLGMAVA
 KRKIALPQTGESESYYPALLGLMLGLTPIPIPKINN"
 complement(3663..5324)
 /gene="recN"
 /note="SpyM3_1151"
 complement(3663..5324)
 /gene="recN"
 /note="best blastp hit: gb|AAL98084.1| (AE010066) putative
 DNA repair and genetic recombination protein
 [Streptococcus pyogenes MGAS8232]"
 /codon_start=1
 /transl_table=11
 /product="putative DNA repair and genetic recombination
 protein"
 /protein_id="AAM79758.1"
 /db_xref="GI:21904890"
 /translation="MLLEISIKNFALIDEISLNFENGMTVLTGTGAGKSIITIDNMN
 MLCARASTEVIRRGANKAEIGFVSVDATPELVACLESQGIAMEEBELIIRDFIFANGR
 SVSRINGQMVNLATLKQVGLVDIHGHQDQELMRPOLHQIILDAFGKAPFALKEN
 YQLIFDYSKSLRQVIDKQNEKEHKDRIDMLAQIAEIAEAAALSAGEDDRLNQERDR
 LMHKQIADTLTNAVYVLDNDPSSLSNRSSNMNLSLIEQDFSETYKGMSTSISEAYI
 LLEVSQKLSDTIDQLDFDGRQLEIETFRDLTNSLTRYGGNDVLDYDNIKEY
 QVLTGDDLSGDSLESELKOLVAAATNLSERRHQALQAEIAEKELKELMEKA
 DKVHPTTSKFNDRGNESLEFYISTNPGEKFLPVKVASGGELSLMLAIKAAISRKE
 DKTSTIVEDVDTGCVGRVAAQIAQIKYIKGRHGQVIAISHLPOVIALADYQYFISKE
 KEESTVSKVRLTPEERVEIASMIAGTDMTQAAALTOARELLAKH"
 complement(5346..5816)

gene
 CDS
 gene
 CDS
 gene
 CDS
 gene
 CDS
 gene
 CDS

```

/codon_start=1
/transl_table=11
/product="glucose kinase"
/protein_id="AAL98113.1"
/db_xref="GI:19748683"
/transl_table="MSQKLGIDLGTTIKFGILTAAGEVOEKWAIETNILEGGKHIV
PDIVASIKHRLDLYGLSSADFGVIGMGSPGAVDRDNTVTGAFNLNWKETQEVGSVVE
KEIGIPPAIDANDANVAALGERWVGAGENNPDVFMFTLGTGVGGGIIADGNLIHGVAGA
GGEGHIMVPENGCFACGSHGCLETVASATGVVKVARLLAEAYEGSDSAIKRAIDNG
EYTSKIDIFMAAEGDSFADSVVEKVGYYILGLASANISNLINPDSDSVVIGGVSAAGEF
LRGRIEIVFTFPQVRYSTKIKIAELGNDAGIIGAASLAROFITK"
/complement(3981..4184)
/gene="spym18.1547"
/complement(3981..4184)
/gene="spym18.1547"
/notes="best blastp match gb|AAK34324.1| (AE006586)
conserved hypothetical protein [Streptococcus pyogenes M1
GAS]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAL98114.1"
/db_xref="GI:19748684"
/transl_table="MKTLYDVQQLKNFGIFVYLGKRLYDIEMMKIELQRLYDSGLLD
KRDYLNALILRREHRLKEKEG"
/complement(4326..4853)
/gene="spym18.1548"
/complement(4326..4853)
/gene="spym18.1548"
/notes="best blastp match gb|AAK34325.1| (AE006586)
putative peroxide resistance protein [Streptococcus
pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative peroxide resistance protein"
/protein_id="AAL98115.1"
/db_xref="GI:19748685"
/transl_table="MTNLTVENIYASVTHNISKKEASNEKTRAVLNQAVDLSSAAS
IVHQVHWYMRGPGFLYLPKMDLSDLSNLDMSERLITIGCAPYSTLAEFSSKSHK
LDBAKGYDKTVAQHGLARLYVEVYLLSSLLYQVGLDITDEBGADTNDLFTAATEAK
TIWMLQNERQGGPAL"
/5036..5706
/gene="spym18.1549"
/5036..5706
/gene="spym18.1549"
/notes="best blastp match gb|AAK34326.1| (AE006586)
hypothetical protein [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAL98116.1"
/db_xref="GI:19748686"
/transl_table="MISFRINKYDSFSCISGILLTHLFVCLLLYFPDNCQEQLSKQT
VIVCDQTFARVRFQNFHTISQKIFFSLEISLIIILWASVHWVWSASYCYLLILFS
LRFSLDFWRQSEYPTLWLFPSVSLFLFYSINYLSELLGLLGLAHLRPFSGAGDFF
YLASLAIVLDLTSLIWLQSLASLAGITACLLGKTRIPFIPYLSFGLFWTLVLEHF"
/complement(5788..6867)
/gene="spym18.1550"
/complement(5788..6867)
/gene="spym18.1550"
/notes="best blastp match gb|AAK34327.1| (AE006586)
conserved hypothetical protein [Streptococcus pyogenes M1
GAS]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAL98117.1"
/db_xref="GI:19748687"
/transl_table="MKPSIYSLTRDELIAWVERGOKQPRATOIWMKLYKKRVQSFEE
MTNISKDFVILNDSFCVNPDKORVQESADGTVKYLFELPDGMLLETVMRQHYGHS
VCVITQVCNIGCTFCASGLIKKORVQESADGTVAQIMLVQRYEDDRKQGERVSVVWM
GIEPFDNYKNVCMCFURVINDNGLAIGARHITVSTVSGLAHKIRDFANEGVQVNLAVS
LHAPNDRSSIMRVNRSFPLEKLFSAIEYYIEKTNRRVTFEYIMLNEVNDISIKQAE

```

```

LADLTITIRKLSYVNLIPYNPVSEHDOYSRSPKERVLAFLVDVLKKGVCVVRQEHGT
DIDAACGOLRSKTKDKREKVATTK"
/complement(6871..7515)
/gene="spym18.1551"
/complement(6871..7515)
/gene="spym18.1551"
/notes="best blastp match gb|AAK34328.1| (AE006586)
hypothetical protein [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAL98118.1"
/db_xref="GI:19748688"
/transl_table="MDFSERCOMKKEISPEWYVKNKPPGPKFIHFEEQVKAEGIDLLI
LEDVKNAPDTSFGQRYTEVLLKYDYIVGDNQELRLKGFYKSDSDIKKTNRLSRLE
DYIKFCNFCATFVLENLHPQDIIKFEERQPRKKSPKSRNRRKPNYNQOPATPK
SKKRVSKERQENQAPTQSKRRSNTHKKEKSRNQTSQLNTKISHFIIRKKDK"
/complement(7939..8916)
/gene="spym18.1552"
/complement(7939..8916)
/gene="spym18.1552"
/notes="best blastp match gb|AAK34329.1| (AE006586)
putative ribose transport operon repressor [Streptococcus
pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative ribose transport operon repressor"
/protein_id="AAL98119.1"
/db_xref="GI:19748689"
/transl_table="MTVIKQVAAEAGVSRVSYISQKGYSDDRHKIKAAIAKILH
LEEVYVHQLSKTKNOLVLLDPISNPPFRLARGAEYLYKKGVYRMLGNISDSEA
LYEYVAVLQSSNAAGIITTHDFTKRYPTLAIPVVVDRVDQETQGVFSDNRAGGL
AAQTVQAGAKEVLLIRGLDNLNENEFESFYLQKODVTMYVCDSDNFDEFESIQ
LEASYNLKCYPTIDSIAPSDIHAIYIHELHSQGGKIPQDVQIIGYDDILMSQFIYP

```

Alignment Scores:

Pred. No.:	1.05e-42	Length:	11465
Score:	557.50	Matches:	107
Percent Similarity:	71.98%	Conservative:	42
Best Local Similarity:	51.69%	Mismatches:	41
Query Match:	55.25%	Indels:	17
DB:	1	Gaps:	4

US-09-868-352-23 (1-192) x AE010069 (1-11465)

Qy 1 MetArgLysGluValThrProGluMetLeuAsnTyrAsnLysTyrProGlyProGlnPhe 20

Db 7491 ATGAAAAAGAAATTCGCCAGAAATGTACAACTATATAAATTTCCAGGTCCTCCAAATTT 7432

Qy 21 ILeHisPheGluAsnIleValLysSerAspAspIleGluPheGlnLeuValIleAsnGlu 40

Db 7431 ATTCATTTTGAAGACCAAGTTAAGCTCAAGGCATTCATTGTTACTCTTAGAAGATGTT 7372

Qy 41 LysSerAlaPheAspValThrValPheGlyGlnArgPheSerGluIleLeuLeuLysTyr 60

Db 7371 AAGAACGCTTTTGATACGACTAGCTTTGGTCAACGTTATACAGAAGTCTCCTAAAGTAT 7312

Qy 61 AspPheIleValGlyAspTrpGlyAsnGluIleLeuArgLeuArgGlyPheTyrLysAsp 80

Db 7311 GACTATATTGTGGCGATTGGGAAATGAACAGCTTCGCTTAAGAGGTTTTTATAAGGAT 7252

Qy 81 AlaSerThrIleArgAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyr 100

Db 7251 AGTGATGATATCAAGAAACCAATCGCATCTCAGCTTTAGAGGATATATTAAGAATTT 7192

Qy 101 CysAsnPheGlyCysAlaTyrPheValIleGluAsnProAsnProArgAspIleLysPhe 120

Db 7191 TGCAATTTTGGTCTGCTTATTTTGTAGAAAAATCTTCACACAGATATTAAATTT 7132

Qy 121 AspAspGluArgProHisLysArgLysSer---ArgSerLysSerGln----- 136

Db 7131 GAAGAGGCGCCCAACCAACGAGAAATAACCTAAATCAAAATCAATCGTCGCAAG 7072


```

/db_xref="GI:13622615"
/translation="MTNLRNDIRNVAIIAHVDHCKTTLVDELLKQSHTLDERKELOER
AMSDNLEKERGITIILAKNTAVAYNDVRINIMDTPGHADEFGGEVERIMKWMDGVVLV
DAYEGTPTOTREVLKALKALQNLIPIVVANKIDKPSARPAEYVDEVELEFELGADDEQ
LEPFGVAGISGTSLSDDPADQDHTMAPIEDTIIIDHIPAPVDNDDEPLQFOVSLLD
YNDVFGRIAGNRTVGVKQDOVTLKSLDGTTKNFRVTKLFGFFGLERREGEAKAG
DIAVAGMEDIFVGETITPTDCVEALPIILRIDETLQMTFLVNNSPFAGREKWIISR
KVBERLAELODVSURVPTDSPDKWTVSGRCELHLSILIEPMRREGYELQVSRPEV
ILKEIDVCEPEPERVIDTPEYOCALIOSLSERKGDMDMDQMGVNGOTRLILFLIPA
RLGLIGSTELSTRGYGINHTFDQYLPVQGEIGRGALVSTENKATYISIMR
IEERGIFVNGTEVIEGMIVGNSDNDLVNITAKQMTNVRSA TKDQTAIVKTPR
ITLESLEFLNDEMEYTPESIRKQILNKAARDKANKKKKSAE"
gene
CDS
/notes="Best Blastp hit = dbj|BAA12549.1| (D84432)
YqHl [Bacillus subtilis] >gi|1303893|dbj|BAA12549.1| (D84432)
YqHl [Bacillus subtilis] >gi|2634888|emb|CAB14385.1|
(299116) similar to hypothetical proteins [Bacillus
subtilis]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK34322.1"
/db_xref="GI:13622616"
/translation="MSPIYLILMLLVGIVGYTWNVYFSEKMAKQVDNETKDVWVRQ
GLQIDUREAPAFKTHILGARNFPAAQFDAAIKGLRKDPVLIIYENRPOYRPAVVK
LKAGFEDYVVLKGDYDWDGKVKQT"
complement(3046..4017)
gene
CDS
/notes="SpY1529"
complement(3046..4017)
/genes="gick"
/EC_number="2.7.1.2"
/notes="Best Blastp hit = pir|T44787 glucokinase (EC
2.7.1.2) [imported] - Bacillus megaterium
>gi|2274967|emb|CAA03848.1| (AJ000005) glucose kinase
[Bacillus megaterium]"
/codon_start=1
/transl_table=11
/product="glucose kinase"
/protein_id="AAK34323.1"
/db_xref="GI:13622617"
/translation="MSQKLLGIDLGTTIKFGLTAAGEVOEKWAIETNILEGKHIH
PDIIASIKRDLIDYGLUSSADFGIGMSPGAVDRDNTVTGAFNLNWKETQEVGSVVE
KELGTPAIDNVAALGERWVAGENPNPDVFMPTLGTGVGGIITDGNLHGVAGA
GGEIGHMIVEPENGFACGSHGCLFETVATGVVAVARLLAEAYEGDSAIAKAIDNG
EGVTSKDIEMAAEAGDSFADSVVEKGYVYGLASANTSNLNPDSVVIIGGVSAAGEF
LRSRIKVFVTFEFPQVRYSTKIKIAELNDAGIIGAAASLARQFITK"
complement(4022..4225)
gene
CDS
/genes="SpY1530"
complement(4022..4225)
/genes="SpY1530"
/notes="Best Blastp hit = emb|CAA03847.1| (AJ000005) orf1
[Bacillus megaterium]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK34324.1"
/db_xref="GI:13622618"
/translation="MKTLYDVQOLLKNGFIVYLGKRLYDIEMMKTELQRLYDSGLLD
KRDYLAELILREHRELEKEG"
complement(4367..4894)
gene
CDS
/genes="dpr"
complement(4367..4894)
/notes="Best Blastp hit = gb|AAG33871.1|AF319974_1
(AF319974) Dpr [Streptococcus suis]"
/codon_start=1

```

```

/transl_table=11
/product="putative peroxide resistance protein"
/protein_id="AAK34325.1"
/db_xref="GI:13622619"
/translation="MTNTLVENIYASVTHNISKEASKEKTRKAVLNQAVADUSVAAS
IVHQHWYMRGFGFLYHPKMDLSDLANDEMGERLITIGGAPYSTLAEFSEKISK
LDEAKGTQYKTAQHLARLVEVYLYSSLYQVGLDITDEGDAGTNDLFTAAKTEAK
TIWMLQAEKQGPAL"
5122..5748
/genes="SpY1532"
5122..5748
/genes="SpY1532"
/notes="Best Blastp hit = dbj|BAA96471.1| (AB036428) type
IV prepilin peptidase homologue [Streptococcus mutans]"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK34326.1"
/db_xref="GI:13622620"
/translation="MIAFFSICSGILLTHFYCLLLYFPDNGOQLSKOTVIYCDOLT
FAKRVPRQNFHITTSOKIFFSLEIGLIIILWASVMWVSASVCYLLLSLFLSFDW
RQGEPPFILWFSFVSLFYSINYLISLILLGLLAHLRPPSIGAGDFYFLASLALV
LDLTSUWLIQIASLAGITACLLLGKIRPIPIYLSFGFLVIVLLEHF"
complement(5830..6909)
gene
CDS
/genes="SpY1533"
complement(5830..6909)
/genes="SpY1533"
/notes="Best Blastp hit = sp|O34617|YLON_BACSU HYPOTHETICAL
14.6 KDA PROTEIN IN FWT-SPOVM INTERGENIC REGION
>gi|7448949|pir|F69878 conserved hypothetical protein
YLON - Bacillus subtilis >gi|2337804|emb|CAA74265.1|
(Y13937) YLON protein [Bacillus subtilis]
>gi|2633947|emb|CAB13448.1| (299112) similar to
hypothetical proteins [Bacillus subtilis]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK34327.1"
/db_xref="GI:13622621"
/translation="MKPSIYSLTRDELIAWVERGOKQFRATQIWNWLYKKRVQSFE
MNISKDFVILNDSFVAPLQKRVVQESADGTVYLFELPDGMLLETVMRHOYHGS
VCVTTQVGCNIGTCFASGLIKKQDLNSETAQIMLVOKYFDDRKQGERVSHVVVM
GICEPFDYKNVCMFLRVINDNGLAIGARHITVSTGLAHKIRDFANEGVQVNLAVS
LHAPNDRLSIRMVNRSPFLKFLSAIEYIEKTRNRTVEYIMLNEVNDISKQAE
LADLTKTIRKLSVYVNLIPYNPVSEHDOYSRSPKRVLFADVLKKNVGVVQVQEHGT
DIDAGQLRSKTMKDKREKATK"
complement(6913..7557)
gene
CDS
/genes="SpY1534"
complement(6913..7557)
/genes="SpY1534"
/notes="Best Blastp hit = pir|F70023 hypothetical protein
yutD - Bacillus subtilis >gi|2635728|emb|CAB15221.1|
(299120) yutD [Bacillus subtilis]"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK34328.1"
/db_xref="GI:13622622"
/translation="MDFSERCOMKKEISPENYNNKFPCKFIHFPEQVKAEGIDLLL
LDVKNAPDTSFGQRYTEVLLYDIYDVGWNEQURLKGFYKDDSDIKKTNRISLE
DYIKFECNFCGAYFVLENLHJPDQDIKEFVEQKSKSPKSNRRPNYSQQPAPTK
SKSKRASKEQENQAFTSQKRRSNTKHKEKSKRNOTSOLNTKISHFIIRKKDK"
complement(7981..8958)
gene
CDS
/genes="SpY1535"
complement(7981..8958)
/genes="SpY1535"
/notes="Best Blastp hit = pir|I40462 repressor of ribose
transport operon - Bacillus subtilis
>gi|580917|emb|CAA81048.1| (225798) repressor of ribose

```

Alignment Scores:
Pred. No.: 11507
Score: 558.50
Matches: 107

AX026665
LOCUS AX026665 579 bp DNA linear PAT 16-SEP-2000
DEFINITION Sequence 22 from Patent WO037646.
ACCESSION AX026665
VERSION AX026665.1 GI:10187839
KEYWORDS Streptococcus sp. 'group B';
SOURCE Streptococcus sp. 'group B';
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 579)
AUTHORS Lane, J.D., Santangelo, J.D., Dougan, G., Everest, P., Feldman, R., Dobson, R.J., Hughes, M.J., Moore, J.C. and Wilson, R.K.
TITLE Genes and proteins, and their use
JOURNAL Patent: WO 0037646-A 22 29-JUN-2000;
LANE JONATHAN DOUGLAS (GB); SANTANGELO JOSEPH DAVID (GB); DOUGAN GORDON (GB); EVEREST PAUL (GB); FELDMAN ROBERT (GB); DOBSON RICHARD JAMES (GB); HUGHES MARTIN JOHN GLENTON (GB); MICROSCIENCE LTD (GB); MOORE JOANNE CHRISTINE (GB); WILSON REBECCA KERRY (GB)
FEATURES
source 1..579
/organism="Streptococcus sp. 'group B'"
/db_xref="taxon:1319"
CDS 1..579
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC09097.1"
/db_xref="GI:10187840"
TRANSLATION="MRKEVTPMLNKNYPGQFIHFENIVKSDDEIFOLVINEKSAF
DVTVGQRFSEILLKVDYFIVGDMGNEQLRGLFKDASPIRKNRSIRSLDEYIKVCN
FCCAVFLNPNPRDIKEDDERPHKRRKRSKSKSQTNRNRSQSNANAHFTSKRR
KDKRQERHIKEQDKWNTSAKHLFLVRKN"
BASE COUNT 213 a 85 c 109 g 172 t
ORIGIN
Alignment Scores:
Pred. No.: 3 12e-87 Length: 579
Score: 1009.00 Matches: 192
Percent Similarity: 100.00 Conservative: 0
Best Local Similarity: 100.00 Mismatches: 0
Query Match: 100.00 Indels: 0
DB: 6 Gaps: 0
US-09-868-352-23 (1-192) x AX026665 (1-579)
QY 1 MetArgLysGluValThrProGluMetLeuAsnTyrAsnLysTyrProGlyProGlnPhe 20
Db 1 ATCGGAAAAGAGTGCACACAGAGATGCTTAACATAATAAGTATCTCGCCACAGTTT 60
QY 21 IleHisPheGluAsnIleValLysSerAspAspIleGluPheGlnLeuValIleAsnGlu 40
Db 61 ATTCACCTTTGAAAATATCGTTAAAGATGATGATATTCAATTTTCAACTTGTATTATGAA 120
QY 41 LysSerAlaPheAspValThrValPheGlyGlnArgPheSerGluIleLeuLysTyr 60
Db 121 AAATCAGCTTTTCATGTGACTGCTTTGGACACAGTTTCTCAGAGATTTATTAAATAT 180
QY 61 AspPheIleValGlyAspTrpGlyAsnGluGlnLeuArgLeuArgGlyPheTyrLysAsp 80
Db 181 GATTTTATCGTTCGGCATTGGGGTACGAGCAGCTTGAGGCTAAGAGCCTTTTACAAAGAT 240
QY 81 AlaSerThrIleArgLysAsnSerArgIleSerArgLeuGluAspTyrIleLysGluTyr 100
Db 241 GCTAGTACCATTTAGAAAAATACCGCGATTTTCACGTTTAGAGATTATATTAAAGAGTAT 300
QY 101 CysAsnPheGlyCysAlaTyrPheValLeuGluAsnProAsnProArgAspIleLysPhe 120
Db 301 TGTAACTTTGGTTGCTTATTTTGTGTGGAGAATCCCAATCCTAGAGATATTAAATTT 360
QY 121 AspAspGluArgProHisLysArgArgLysSerArgSerLysSerGlnSerLysSer 140
Db 361 GATGATGAAGACCTCATAGCGCTCGTAAGTCAAGATCCCAATCAACAATCATCAAGTCA 420

QY 141 GlnThrArgAsnAsnArgSerGlnSerAsnAlaAsnAlaHisPheThrSerLysArg 160
Db 421 CAAACTAGAAATAATCGTCCAGTCAAAATGCCAATGCTCATTTTACAAAGTAAAGCGT 480
QY 161 LysAspThrLysArgArgGlnGluArgHisIleLysGluGlnAspLysGluMetThr 180
Db 481 AAAGACACAAACGCCGTCGAAGAACGTCATATTAAAGAGAGCAAGATAAGGAATGACC 540
QY 181 SerAlaLysGlnHisLeuLeuPheValArgLysAsn 192
Db 541 TCTGCAACAGCAGCATTTGTTATTTCGTAAGAAAAAT 576
RESULT 2
AE006586/11 11507 bp DNA linear BCT 01-JUN-2001
LOCUS Streptococcus pyogenes M1 GAS strain SF370, section 115 of 167 of
DEFINITION the complete genome.
ACCESSION AE006586 AE004092
VERSION AE006586.1 GI:13622613
KEYWORDS Streptococcus pyogenes M1 GAS.
SOURCE Streptococcus pyogenes M1 GAS
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 11507)
AUTHORS Ferretti, J.J., McShan, W.M., Adjic, D., Savic, D., Savic, G., Lyon, K., Primeaux, C., Sezate, S., Surorov, A.N., Kenton, S., Lai, H., Lin, S., Qian, Y., Jia, H.G., Najjar, F.Z., Ren, Q., Zhu, H., Song, L., White, J., Yuan, X., Clifton, S.W., Roe, B.A. and McLaughlin, R.E.
TITLE Complete genome sequence of an M1 strain of Streptococcus pyogenes
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (8), 4658-4663 (2001)
MEDLINE 21192684
PUBMED 11296296
REFERENCE 2 (bases 1 to 11507)
AUTHORS Ferretti, J.J., McShan, W.M., Adjic, D., Savic, D., Savic, G., Lyon, K., Primeaux, C., Sezate, S., Surorov, A.N., Kenton, S., Lai, H., Lin, S., Qian, Y., Jia, H.G., Najjar, F.Z., Ren, Q., Zhu, H., Song, L., White, J., Yuan, X., Clifton, S.W., Roe, B.A. and McLaughlin, R.E.
TITLE Direct Submission
JOURNAL Submitted (10-APR-2001) Department of Microbiology and Immunology,
University of Oklahoma Health Sciences Center, 940 SL Young Blvd,
Oklahoma City, OK 73104, USA
FEATURES
Location/Qualifiers
1..11507
/organism="Streptococcus pyogenes M1 GAS"
/strain="SF370"
/serotype="M1"
/db_xref="taxon:160490"
complement(250..501)
/gene="SPY1526"
complement(250..501)
/gene="SPY1526"
/note="Best Blastp hit = no blast hits"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK34320.1"
/db_xref="GI:13622614"
/translation="MFYLIMAVSIISYLYMAPKSVRNLTGLMGLVALLIVLAGL
SFIKIMQTPPEFICGMVVALGYFALKDVRKTKPRVK"
complement(623..2464)
/gene="SPY1527"
complement(623..2464)
/gene="SPY1527"
/note="Best Blastp hit = dbj|BAB06351.1| (AP001516)
GTP-binding protein TypA/BipA (tyrosine phosphorylated
protein A) (Bacillus halodurans]"
/codon_start=1
/transl_table=11
/product="putative GTP-binding protein TypA/BipA (tyrosine
phosphorylated protein A)"
/protein_id="AAK34321.1"

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 25, 2003, 09:38:47 ; Search time 2693 Seconds
(without alignments)
2074.912 Million cell updates/sec

Title: US-09-868-352-23
Perfect score: 1009
Sequence: 1 MRKEVTPEMLNKNYPGPQF.....EEQDKEMTSKQHLFVRKN 192

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_plus_p2n_model -DEV=xlh
-Q/cgcn2_1/USPRO_spool/US09868352/runat_14022003_153516_4220/app_query.fasta_1.391
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=40 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09868352 -CGCN_1.1_2425 -runat_14022003_153516_4220 -NCPU=6 -ICPU=3
-NO_XLPAY -NO_MAP -LARGESUB -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:*

1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hun:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rdd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1009	100.0	579	6	AX026665 Sequence
C 2	558.5	55.4	11507	1	AE006586 Streptoco
C 3	557.5	55.3	11465	1	AE010069 Streptoco
C 4	557.5	55.3	50463	1	AE014159 Streptoco
5	452.5	44.8	10029	1	AE008445 Streptoco
C 6	452.5	44.8	147006	2	SPNEU1904
C 7	438	43.4	5558	6	BD003783 Polynucle
8	438	43.4	10029	1	AE007384 Streptoco
9	331.5	32.9	10098	1	AE006255 Lactococc
C 10	255.5	25.3	3054	1	AF270167 Staphyloc
C 11	255.5	25.3	3054	6	AX145485 Sequence
12	249.5	24.7	290150	1	AP004824 Staphyloc
13	249.5	24.7	299050	1	AP003131 Staphyloc
14	249.5	24.7	348527	1	AP003360 Staphyloc
C 15	248	24.6	248050	1	AL596172 Listeria
C 16	248	24.6	349980	6	AX413018 Sequence
C 17	248	24.6	349980	6	AX417048 Sequence
18	243.5	24.1	390	6	AX142599 Sequence
C 19	240.5	23.8	324050	1	AL591983 Listeria
20	239.5	23.7	1234	6	AX414540 Sequence
21	239.5	23.7	2690	6	AX416642 Sequence
C 22	234.5	23.2	303250	1	AP001518 Bacillus
C 23	232.5	23.0	285	6	AX439521 Sequence
C 24	232.5	23.0	217420	1	BSUB0017 Bacillus su
25	222.5	22.1	273	6	AX434730 Sequence
26	113.5	11.2	6145	6	E26748 Structure a
27	113	11.2	179908	10	AC125344
28	109.5	10.9	4615	8	SCYGL133W
29	109.5	10.9	14607	8	SCX11
30	108.5	10.8	2204	10	BC019437
31	108.5	10.8	127044	2	AC098505 Rattus no
32	108	10.7	215455	2	AC083856 Mus muscu
33	107.5	10.7	3130	10	BC026944 Mus muscu
34	107.5	10.7	207814	10	AL590994 Mouse DNA
35	107	10.6	218677	2	AC109286 Mus muscu
36	106.5	10.6	25117	2	AC116102 Dictyoste
37	106.5	10.6	237855	2	AL391558 Homo sapi
C 38	104.5	10.4	100269	8	ATF18022 Arabidops
C 39	104.5	10.4	132934	2	AC120073 Rattus no
C 40	104.5	10.4	168111	2	AC096061 Rattus no
C 41	104	10.3	205764	2	AC079478 Mus muscu
C 42	103.5	10.3	191914	2	AC117251 Mus muscu
43	103.5	10.3	214403	2	AC126270 Mus muscu
44	102.5	10.2	69894	2	AC123604 Mus muscu
45	102.5	10.2	115049	2	AC073746 Mus muscu

ALIGNMENTS

RESULT 1

; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-09-984-198-150

Query Match 3.1%; Score 6; DB 10; Length 258;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 KEVTP 8
Db 68 KEVTP 73

RESULT 40
US-10-024-579-12
; Sequence 12, Application US/10024579
; Patent No. US20020119522A1
; GENERAL INFORMATION:
; APPLICANT: Friddle, Carl Johan
; APPLICANT: Gerhardt, Brenda
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20020119522A1el Human Ion Channel-Related Proteins
; FILE REFERENCE: LEX-0274-USA
; CURRENT APPLICATION NUMBER: US/10/024,579
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: US 60/258,595
; PRIOR FILING DATE: 2000-12-28
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 264
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-024-579-12

Query Match 3.1%; Score 6; DB 12; Length 264;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 88 SRISRL 93
Db 80 SRISRL 85

Search completed: February 26, 2003, 10:05:16
Job time : 20 secs

Db 138 VKSDDI 143

```
RESULT 33
US-09-774-639-155
; Sequence 155, Application US/09774639
; Publication No. US20030003555A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 90 Human Secreted Proteins
; FILE REFERENCE: P2013P1
; CURRENT APPLICATION NUMBER: US/09/774,639
; CURRENT FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/244,112
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-02-04
; NUMBER OF SEQ ID NOS: 371
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 155
; LENGTH: 227
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-774-639-155
```

```
Query Match 3.1%; Score 6; DB 9; Length 227;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 RKSRSK 134
|||||
Db 66 RKSRSK 71
```

```
RESULT 34
US-09-738-626-4680
; Sequence 4680, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENO, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 4680
; LENGTH: 235
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-4680
```

```
Query Match 3.1%; Score 6; DB 9; Length 235;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 88 SRISRL 93
|||||
Db 213 SRISRL 218
```

```
RESULT 35
US-09-738-626-6304
; Sequence 6304, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENO, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6304
; LENGTH: 236
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6304
```

```
Query Match 3.1%; Score 6; DB 9; Length 236;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 DKEMTS 181
|||||
Db 70 DKEMTS 75
```

```
RESULT 36
US-10-024-579-16
; Sequence 16, Application US/10024579
; Patent No. US20020119522A1
; GENERAL INFORMATION:
; APPLICANT: Fiddle, Carl Johan
; APPLICANT: Gerhardt, Brenda
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20020119522A1 Human Ion Channel-Related Proteins
; TITLE OF INVENTION: and Polynucleotides Encoding the Same
; FILE REFERENCE: LEX-0274-USA
; CURRENT APPLICATION NUMBER: US/10/024,579
; CURRENT FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: US 60/258,595
; PRIOR FILING DATE: 2000-12-28
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 238
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-024-579-16
```

```
Query Match 3.1%; Score 6; DB 12; Length 238;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 88 SRSRL 93
|||||
```

; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA: 09/106,582
; APPLICATION NUMBER: 09/106,582
; FILING DATE: 29-JUN-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.439C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-682-6031
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 196 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORGANISM: Ehrlichia
; US-09-159-469-32

Query Match 3.1%; Score 6; DB 10; Length 196;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 VKSDDI 32
|||||
Db 138 VKSDDI 143

RESULT 30
US-09-159-469-58
; Sequence 58, Application US/09159469
; Patent No. US20020064535A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
; TITLE OF INVENTION: THERAPY OF EHRlichIA INFECTION
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/159,469
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/106,582
; FILING DATE: 29-JUN-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.439C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-682-6031
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 196 amino acids
; TYPE: amino acid

; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-159-469-58

Query Match 3.1%; Score 6; DB 10; Length 196;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 VKSDDI 32
|||||
Db 138 VKSDDI 143

RESULT 31
US-09-798-042-32
; Sequence 32, Application US/09798042
; Patent No. US20020068343A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; TITLE OF INVENTION: AND TREATMENT OF EHRlichIA INFECTION
; FILE REFERENCE: 210121.439C7
; CURRENT APPLICATION NUMBER: US/09/798,042
; CURRENT FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 108
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 196
; TYPE: PRT
; ORGANISM: Ehrlichia sp.
; US-09-798-042-32

Query Match 3.1%; Score 6; DB 10; Length 196;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 VKSDDI 32
|||||
Db 138 VKSDDI 143

RESULT 32
US-09-798-042-58
; Sequence 58, Application US/09798042
; Patent No. US20020068343A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; TITLE OF INVENTION: AND TREATMENT OF EHRlichIA INFECTION
; FILE REFERENCE: 210121.439C7
; CURRENT APPLICATION NUMBER: US/09/798,042
; CURRENT FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 108
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 58
; LENGTH: 196
; TYPE: PRT
; ORGANISM: Ehrlichia sp.
; US-09-798-042-58

Query Match 3.1%; Score 6; DB 10; Length 196;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 VKSDDI 32
|||||

```
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 33652
; LENGTH: 138
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO 285090.10
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 32
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1e+02
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 3.3e+02
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 56
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.2e+02
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2e+02
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 26
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1e+02
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 90
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.2e+02
; OTHER INFORMATION: EST_HUMAN HIT: BE742115.1, EVALUO 2.00e-49
; OTHER INFORMATION: EST_HUMAN HIT: BE299019.1, EVALUO 3.00e-49
; OTHER INFORMATION: SWISSPROT HIT: O43683, EVALUO 5.60e-01
; US-09-864-761-33652

Query Match
Best Local Similarity 100.0%; Pred. No. 1e+02; Length 138;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 DVTVEG 50
DB 43 DVTVEG 48
```

```
RESULT 27
US-09-764-864-1566
; Sequence 1566, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; PRIOR APPLICATION DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1566
; LENGTH: 146
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-764-864-1566

Query Match
Best Local Similarity 100.0%; Pred. No. 1.1e+02; Length 146;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 127 KRRKSR 132
DB 32 KRRKSR 37
|||||

RESULT 28
US-09-925-300-1731
; Sequence 1731, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben,
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1731
; LENGTH: 156
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (38)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (134)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (143)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; US-09-925-300-1731

Query Match
Best Local Similarity 100.0%; Pred. No. 1.1e+02; Length 156;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 KEVTPE 8
DB 106 KEVTPE 111
|||||
```

```
RESULT 29
US-09-159-469-32
; Sequence 32, Application US/09159469
; Patent No. US20020064535A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
; THERAPY OF EHRlichia INFECTION
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/159,469
```

; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1527
; LENGTH: 75
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-796-692-1527

Query Match 3.1%; Score 6; DB 9; Length 75;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 132 RSKSQS 137
| | | | |
Db 67 RSKSQS 72

RESULT 24
US-09-764-869-945
; Sequence 945, Application US/09764869
; Patent No. US20020061521A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007
; CURRENT APPLICATION NUMBER: US/09/764,869
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2442
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 945
; LENGTH: 88
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (2)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (29)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-869-945

Query Match 3.1%; Score 6; DB 10; Length 88;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 145 NRSQSN 150
| | | | |
Db 75 NRSQSN 80

RESULT 25
US-09-925-299-1067
; Sequence 1067, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12

; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1067
; LENGTH: 98
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (4)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (8)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (27)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (37)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (68)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (69)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (73)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (83)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-1067

Query Match 3.1%; Score 6; DB 10; Length 98;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 3 KEVTPE 8
| | | | |
Db 75 KEVTPE 80

RESULT 26
US-09-864-761-33f52
; Sequence 33652, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Pern, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomlca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669

GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aemica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 35132
LENGTH: 49
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC009955.1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 3.1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.5
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.5
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.4
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.4
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.2
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.1
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.7
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.9
US-09-864-761-35132

Query Match 3.1%; Score 6; DB 10; Length 49;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 KRROER 169
|||||
DB 24 KRROER 29

RESULT 22
US-09-764-878-190
Sequence 190, Application US/09764878
Patent No. US20020090615A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PA121
CURRENT APPLICATION NUMBER: US/09/764,878
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 428
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 190
LENGTH: 73
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (34)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (35)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (70)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (73)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-878-190

Query Match 3.1%; Score 6; DB 10; Length 73;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 93 LEDYIK 98
|||||
DB 23 LEDYIK 28

RESULT 23
US-09-796-692-1527
Sequence 1527, Application US/09796692
Publication No. US20020198362A1
GENERAL INFORMATION:
APPLICANT: Gaiger, Alexander
APPLICANT: Algate, Paul A.
APPLICANT: Mannion, Jane
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THER
FILE REFERENCE: 2077.001200
CURRENT APPLICATION NUMBER: US/09/796,692
CURRENT FILING DATE: 2001-03-01
PRIOR APPLICATION NUMBER: 60/186,126
PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: 60/190,479
PRIOR FILING DATE: 2000-03-17
PRIOR APPLICATION NUMBER: 60/200,545
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: 60/200,303
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/200,779
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/200,999
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: 60/202,084
PRIOR FILING DATE: 2000-05-04
PRIOR APPLICATION NUMBER: 60/206,201
PRIOR FILING DATE: 2000-05-22
PRIOR APPLICATION NUMBER: 60/218,950

FILE REFERENCE: Acomica-x-1
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/232,366
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 40249
LENGTH: 38
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC009155.3
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 8.3
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 11
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 6.8
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 9.7
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 7.7
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 7.5
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 7.5
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 9.1
OTHER INFORMATION: EST_HUMAN HIT: AUL33018.1, EVALUE 4.00e-05
US-09-864-761-40249

Query Match 3.1%; Score 6; DB 10; Length 38;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 SEILLK 59
DB 29 SEILLK 34

RESULT 20
US-09-864-761-37254
Sequence 37254, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY

FILE REFERENCE: Acomica-x-1
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 37251
LENGTH: 42
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC007314.3
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.2
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.93
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.88
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.95
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.82
OTHER INFORMATION: EST_HUMAN HIT: BE184944.1, EVALUE 4.00e-16
US-09-864-761-37254

Query Match 3.1%; Score 6; DB 10; Length 42;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 KRKSR 132
DB 2 KRKSR 7

RESULT 21
US-09-864-761-35132
Sequence 35132, Application US/09864761
Patent No. US20020048763A1

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 132 RSKSQSS 138
|||||
Db 267 RSKSQSS 273

RESULT 16

US-10-085-027-1
; Sequence 1, Application US/10085027
; Patent No. US20020132759A1
; GENERAL INFORMATION:
; APPLICANT: YAZAKI, YOSHIO
; APPLICANT: ASANO, TOMOICHIRO
; APPLICANT: KUBO, HIDEO
; APPLICANT: KANDA, AKIRA
; TITLE OF INVENTION: REMEDIES FOR DISEASES CAUSED BY INSULIN RESISTANCE
; FILE REFERENCE: 4895-0019-0PCT
; CURRENT APPLICATION NUMBER: US/10/085,027
; CURRENT FILING DATE: 2002-03-01
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: PCT/JP98/04293
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: JP9-263719
; PRIOR FILING DATE: 1997-09-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 1242
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-085-027-1

Query Match 3.6%; Score 7; DB 12; Length 1242;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 132 RSKSQSS 138
|||||
Db 267 RSKSQSS 273

RESULT 17

US-10-038-612-27
; Sequence 27, Application US/10038612
; Patent No. US20020160478A1
; GENERAL INFORMATION:
; APPLICANT: Ben-Sasson, Shmuel A.
; TITLE OF INVENTION: Short Peptides Which Selectively
; TITLE OF INVENTION: Modulate the Activity of Protein Kinases
; FILE REFERENCE: 1242.1029-000 (CMCC-679)
; CURRENT APPLICATION NUMBER: US/10/038,612
; CURRENT FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 09/161,094
; PRIOR FILING DATE: 1998-09-25
; NUMBER OF SEQ ID NOS: 172
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 19
; TYPE: PRT
; ORGANISM: unknown
; FEATURE:
; OTHER INFORMATION: POLO
US-10-038-612-27

Query Match 3.1%; Score 6; DB 9; Length 19;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 126 HKRRKS 131
|||||
Db 13 HKRRKS 18

RESULT 18

US-09-071-838-217
; Sequence 217, Application US/09071838
; Patent No. US20020152501A1
; GENERAL INFORMATION:
; APPLICANT: Fischer, Robert L.
; APPLICANT: Ohad, Nir
; APPLICANT: Kiyosue, Tomohiro
; APPLICANT: Yadegari, Ramin
; APPLICANT: Margossian, Linda
; APPLICANT: Harada, John
; APPLICANT: Goldberg, Robert B.
; TITLE OF INVENTION: Nucleic Acids That Control Seed and
; TITLE OF INVENTION: Fruit Development in Plants
; NUMBER OF SEQUENCES: 324
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,838
; FILING DATE: 01-MAY-1998
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 023070-0861000US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 217:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-071-838-217

Query Match 3.1%; Score 6; DB 10; Length 25;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 86 KNSRIS 91
|||||
Db 9 KNSRIS 14

RESULT 19

US-09-864-761-40249
; Sequence 40249, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04

; APPLICANT: Deutch, Alan H
; APPLICANT: Ghanbari, Hossein A
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS
; FILE REFERENCE: 21486-032 CIP
; CURRENT APPLICATION NUMBER: US/09/859,604
; CURRENT FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/436,184
; PRIOR FILING DATE: 1999-11-08
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1242
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-859-604-5

Query Match 3.6%; Score 7; DB 10; Length 1242;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 RSKSQSS 138
Db 267 RSKSQSS 273

RESULT 12
US-09-903-063-5
; Sequence 5, Application US/09903063
; Patent No. US20020114810A1
; GENERAL INFORMATION:
; APPLICANT: Wands, Jack R.
; APPLICANT: de la Monte, Suzanne M.
; APPLICANT: Ince, Nedim
; APPLICANT: Carlson, Rolf I.
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS
; FILE REFERENCE: 21486-032 DIV3
; CURRENT APPLICATION NUMBER: US/09/903,063
; CURRENT FILING DATE: 2001-10-11
; PRIOR APPLICATION NUMBER: 09/436,184
; PRIOR FILING DATE: 1999-11-08
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1242
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-903-063-5

Query Match 3.6%; Score 7; DB 10; Length 1242;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 RSKSQSS 138
Db 267 RSKSQSS 273

RESULT 13
US-09-903-216-5
; Sequence 5, Application US/09903216
; Patent No. US20020114811A1
; GENERAL INFORMATION:
; APPLICANT: Wands, Jack R.
; APPLICANT: de la Monte, Suzanne M.
; APPLICANT: Ince, Nedim
; APPLICANT: Carlson, Rolf I.
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS
; FILE REFERENCE: 21486-032 DIV2
; CURRENT APPLICATION NUMBER: US/09/903,216
; CURRENT FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: 09/436,184
; PRIOR FILING DATE: 1999-11-08
; NUMBER OF SEQ ID NOS: 9

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1242
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-903-216-5

Query Match 3.6%; Score 7; DB 10; Length 1242;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 RSKSQSS 138
Db 267 RSKSQSS 273

RESULT 14
US-09-903-199-5
; Sequence 5, Application US/09903199
; Patent No. US20020122802A1
; GENERAL INFORMATION:
; APPLICANT: Wands, Jack R.
; APPLICANT: de la Monte, Suzanne M.
; APPLICANT: Ince, Nedim
; APPLICANT: Carlson, Rolf I.
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS
; FILE REFERENCE: 21486-032 DIV4
; CURRENT APPLICATION NUMBER: US/09/903,199
; CURRENT FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: 09/436,184
; PRIOR FILING DATE: 1999-11-08
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1242
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-903-199-5

Query Match 3.6%; Score 7; DB 10; Length 1242;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 RSKSQSS 138
Db 267 RSKSQSS 273

RESULT 15
US-09-903-023-5
; Sequence 5, Application US/09903023
; Patent No. US20020146421A1
; GENERAL INFORMATION:
; APPLICANT: Wands, Jack R.
; APPLICANT: de la Monte, Suzanne M.
; APPLICANT: Ince, Nedim
; APPLICANT: Carlson, Rolf I.
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS
; FILE REFERENCE: 21486-032 DIV1
; CURRENT APPLICATION NUMBER: US/09/903,023
; CURRENT FILING DATE: 2001-10-11
; PRIOR APPLICATION NUMBER: 09/436,184
; PRIOR FILING DATE: 1999-11-08
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1242
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-903-023-5

Query Match 3.6%; Score 7; DB 10; Length 1242;
Best Local Similarity 100.0%; Pred. No. 68;

NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/146,474
FILING DATE: 14-May-2002
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/937,067
FILING DATE: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Lehnhardt, Susan K.
REGISTRATION NUMBER: 33,943
REFERENCE/DOCKET NUMBER: 23647-20018.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 813-5600
TELEFAX: (650) 494-0792
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 317 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-146-474-6

Query Match 3.6%; Score 7; DB 9; Length 317;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 159 KRKDKR 165
DB 236 KRKDKR 242

RESULT 8
US-09-866-582-16
Sequence 16, Application US/09866582
Patent No. US20020127620A1
GENERAL INFORMATION:
APPLICANT: Witman, George B.
APPLICANT: Pazour, Gregory J.
APPLICANT: Rosenbaum, Joel L.
APPLICANT: Cole, Douglas G.
TITLE OF INVENTION: INTRACELLULAR TRANSPORT
FILE REFERENCE: 07917-145001
CURRENT FILING DATE: 2001-05-24
PRIOR FILING DATE: 2000-05-24
NUMBER OF SEQ ID NOS: 50
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 16
LENGTH: 782
TYPE: PRT
ORGANISM: Chlamydomonas reinhardtii
US-09-866-582-16

Query Match 3.6%; Score 7; DB 10; Length 782;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 DWGNEQL 72

DB 769 DWGNEQL 775
RESULT 9
US-09-436-184-5
Sequence 5, Application US/09436184
Publication No. US20030031670A1
GENERAL INFORMATION:
APPLICANT: Wands, Jack R.
APPLICANT: de la Monte, Suzanne M.
APPLICANT: Ince, Nedim
APPLICANT: Carlson, Rolf I.
TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS
FILE REFERENCE: R.I. Hosp. - Malignant Neoplasms
CURRENT APPLICATION NUMBER: US/09/436,184
CURRENT FILING DATE: 1999-11-08
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 5
LENGTH: 1242
TYPE: PRT
ORGANISM: Homo sapiens
US-09-436-184-5

Query Match 3.6%; Score 7; DB 9; Length 1242;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 RSKSQSS 138
DB 267 RSKSQSS 273

RESULT 10
US-09-903-248-5
Sequence 5, Application US/09903248
Patent No. US20020102283A1
GENERAL INFORMATION:
APPLICANT: Wands, Jack R.
APPLICANT: de la Monte, Suzanne M.
APPLICANT: Ince, Nedim
APPLICANT: Carlson, Rolf I.
TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS
FILE REFERENCE: 21486-032 DIV5
CURRENT APPLICATION NUMBER: US/09/903,248
CURRENT FILING DATE: 2001-07-11
PRIOR APPLICATION NUMBER: 09/436,184
PRIOR FILING DATE: 1999-11-08
NUMBER OF SEQ ID NOS: 9
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 5
LENGTH: 1242
TYPE: PRT
ORGANISM: Homo sapiens
US-09-903-248-5

Query Match 3.6%; Score 7; DB 10; Length 1242;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 RSKSQSS 138
DB 267 RSKSQSS 273

RESULT 11
US-09-859-604-5
Sequence 5, Application US/09859604
Patent No. US20020110559A1
GENERAL INFORMATION:
APPLICANT: Wands, Jack R.
APPLICANT: de la Monte, Suzanne M

; CURRENT APPLICATION NUMBER: US/10/085,027
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: 09/508,691
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: PCT/JP98/04293
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: JP9-263719
; PRIOR FILING DATE: 1997-09-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Artificial Sequence
; NAME/KEY: misc feature
; OTHER INFORMATION: Description of Artificial Sequence:synthetic peptide
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (9)..(9)
; OTHER INFORMATION: PHOSPHORYLATION
US-10-085-027-2

Query Match 3.6%; Score 7; DB 12; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 132 RKSQSS 138
|||||
Db 6 RKSQSS 12

RESULT 4

US-10-117-846-19
; Sequence 19, Application US/10117846
; Patent No. US20020168673A1
; GENERAL INFORMATION:
; APPLICANT: Fuller, Margaret T
; APPLICANT: Hales, Karen G.
; APPLICANT: Santel, Ansgar H.
; TITLE OF INVENTION: Mitofusins, Fzo Homologs and Functional
; TITLE OF INVENTION: Derivatives Thereof
; FILE REFERENCE: STAN-063CIP3
; CURRENT APPLICATION NUMBER: US/10/117,846
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: 09/413,285
; PRIOR FILING DATE: 1999-10-06
; PRIOR APPLICATION NUMBER: PCT/US00/27871
; PRIOR FILING DATE: 2000-10-06
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 157
; TYPE: PRT
; ORGANISM: S. cerevisiae
US-10-117-846-19

Query Match 3.6%; Score 7; DB 9; Length 157;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 53 FSEILK 59
|||||
Db 53 FSEILK 59

RESULT 5

US-10-085-027-5
; Sequence 5, Application US/10085027
; Patent No. US20020132759A1
; GENERAL INFORMATION:
; APPLICANT: YAZAKI, YOSHIO
; APPLICANT: ASANO, TOMOICHIRO

; APPLICANT: KUBO, HIDEO
; APPLICANT: KANDA, AKIRA
; TITLE OF INVENTION: REMEDIES FOR DISEASES CAUSED BY INSULIN RESISTANCE
; FILE REFERENCE: 4895-0019-0PCT
; CURRENT APPLICATION NUMBER: US/10/085,027
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: 09/508,691
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: PCT/JP98/04293
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: JP9-263719
; PRIOR FILING DATE: 1997-09-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 159
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-085-027-5

Query Match 3.6%; Score 7; DB 12; Length 159;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 132 RKSQSS 138
|||||
Db 108 RKSQSS 114

RESULT 6

US-09-867-550-1934
; Sequence 1934, Application US/09867550
; Patent No. US20020082206A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Mehrabian, Fuad,
; APPLICANT: Conley, Pamela
; APPLICANT: Law, Debbie
; APPLICANT: Topper, James
; TITLE OF INVENTION: No. US20020082206A1 Polynucleotides from Atherogenic Cells a
; TITLE OF INVENTION: Thereby
; FILE REFERENCE: 21402-013 (Cura-313)
; CURRENT APPLICATION NUMBER: US/09/867,550
; CURRENT FILING DATE: 2001-09-20
; PRIOR APPLICATION NUMBER: USSN 60/208,427
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 2125
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1934
; LENGTH: 200
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-867-550-1934

Query Match 3.6%; Score 7; DB 10; Length 200;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 129 RKSRKS 135
|||||
Db 42 RKSRKS 48

RESULT 7

US-10-146-474-6
; Sequence 6, Application US/10146474
; Publication No. US20030023061A1
; GENERAL INFORMATION:
; APPLICANT: Umansky, Samuil
; APPLICANT: Melkonyan, Hovsep
; TITLE OF INVENTION: A FAMILY OF GENES ENCODING
; APOPTOSIS-RELATED PEPTIDES; PEPTIDES ENCODED THEREBY AND
; METHODS OF USE THEREOF

```

969      5      2.6      272      9      US-09-828-523A-2      Sequence 2, Appli
970      5      2.6      272      9      US-09-738-626-5598      Sequence 5398, Ap
971      5      2.6      274      10     US-09-815-242-12496      Sequence 12496, A
972      5      2.6      275      10     US-09-051-755-5      Sequence 5, Appli
973      5      2.6      276      9      US-09-738-626-4122      Sequence 4122, Ap
974      5      2.6      276      10     US-09-864-761-46690      Sequence 46690, A
975      5      2.6      276      10     US-09-925-300-1169      Sequence 1169, Ap
976      5      2.6      277      10     US-09-925-300-1467      Sequence 1467, Ap
977      5      2.6      278      10     US-09-810-264-28      Sequence 28, Appli
978      5      2.6      278      10     US-09-246-129B-6      Sequence 6, Appli
979      5      2.6      278      10     US-09-872-523-41      Sequence 41, Appli
980      5      2.6      278      10     US-09-899-059-6      Sequence 6, Appli
981      5      2.6      279      8      US-08-971-317A-4      Sequence 4, Appli
982      5      2.6      279      9      US-10-017-910-5      Sequence 5, Appli
983      5      2.6      279      10     US-09-864-761-34389      Sequence 34389, A
984      5      2.6      279      10     US-09-193-663-4      Sequence 4, Appli
985      5      2.6      279      10     US-09-861-451A-28      Sequence 28, Appli
986      5      2.6      279      12     US-10-014-927-19      Sequence 19, Appli
987      5      2.6      280      10     US-09-835-788A-16      Sequence 16, Appli
988      5      2.6      281      8      US-08-971-317A-6      Sequence 6, Appli
989      5      2.6      281      9      US-09-131-237-6      Sequence 6, Appli
990      5      2.6      281      9      US-09-802-669-25      Sequence 25, Appli
991      5      2.6      281      10     US-09-193-663-6      Sequence 6, Appli
992      5      2.6      281      10     US-09-027-287-6      Sequence 6, Appli
993      5      2.6      281      10     US-09-252-656B-6      Sequence 6, Appli
994      5      2.6      281      10     US-09-929-493-6      Sequence 6, Appli
995      5      2.6      281      10     US-09-927-110-1      Sequence 1, Appli
996      5      2.6      281      12     US-10-012-452-13      Sequence 13, Appli
997      5      2.6      282      9      US-10-167-555-2      Sequence 2, Appli
998      5      2.6      282      10     US-09-925-300-1694      Sequence 1694, Ap
999      5      2.6      283      10     US-09-925-302-533      Sequence 533, App
1000     5      2.6      283      10     US-09-925-302-693      Sequence 693, App

```

ALIGNMENTS

```

RESULT 1
US-09-815-242-13414
; Sequence 13414, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13414
; LENGTH: 176

```

```

; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-815-242-13414
Query Match
Best Local Similarity 6.2%; Score 12; DB 10; Length 176;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 69 NEQLRLRGFYKD 80
DB 69 NEQLRLRGFYKD 80
|||||
RESULT 2
US-09-815-242-13567
; Sequence 13567, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13567
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-815-242-13567
Query Match
Best Local Similarity 6.2%; Score 12; DB 10; Length 176;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 69 NEQLRLRGFYKD 80
DB 69 NEQLRLRGFYKD 80
|||||
RESULT 3
US-10-085-027-2
; Sequence 2, Application US/10085027
; Patent No. US20020132759A1
; GENERAL INFORMATION:
; APPLICANT: YAZAKI, YOSHIO
; APPLICANT: ASANO, TOMOICHIRO
; APPLICANT: KUBO, HIDEO
; APPLICANT: KANDA, AKIRA
; TITLE OF INVENTION: REMEDIES FOR DISEASES CAUSED BY INSULIN RESISTANCE
; FILE REFERENCE: 4895-0019-0PCT

```


677	5	2.6	197	10	US-09-925-300-1178	Sequence 1178, Ap	750	5	2.6	220	10	US-09-864-761-33457	Sequence 33457, A
678	5	2.6	198	10	US-09-815-242-12292	Sequence 12292, A	751	5	2.6	221	10	US-09-745-763-152	Sequence 152, Ap
679	5	2.6	199	10	US-09-815-242-13852	Sequence 13852, A	752	5	2.6	221	9	US-09-861-097-22	Sequence 22, Appl
680	5	2.6	199	10	US-09-921-640-6	Sequence 6, Appli	753	5	2.6	221	10	US-09-816-028A-37	Sequence 37, Appl
681	5	2.6	199	10	US-09-764-864-1495	Sequence 1495, Ap	754	5	2.6	221	10	US-09-881-752A-60	Sequence 60, Appl
682	5	2.6	201	10	US-09-840-787-21	Sequence 21, Appl	755	5	2.6	223	10	US-09-893-737-312	Sequence 312, App
683	5	2.6	201	10	US-09-866-987-7	Sequence 7, Appli	756	5	2.6	223	12	US-10-062-254-204	Sequence 204, Appl
684	5	2.6	202	9	US-09-764-868-1018	Sequence 1018, Ap	757	5	2.6	224	10	US-09-731-030A-21	Sequence 21, Appl
685	5	2.6	202	9	US-09-854-133-38	Sequence 38, Appl	758	5	2.6	225	9	US-09-764-868-1102	Sequence 1102, Ap
686	5	2.6	202	10	US-09-738-973-38	Sequence 38, Appl	759	5	2.6	226	9	US-09-736-457-1807	Sequence 1807, Ap
687	5	2.6	202	10	US-09-764-864-1073	Sequence 1073, Ap	760	5	2.6	226	9	US-09-902-941-1807	Sequence 1807, Ap
688	5	2.6	203	10	US-09-925-297-716	Sequence 716, App	761	5	2.6	226	9	US-09-849-626-1807	Sequence 1807, Ap
689	5	2.6	204	9	US-09-738-626-6105	Sequence 6105, Ap	762	5	2.6	226	10	US-09-815-242-10884	Sequence 10884, A
690	5	2.6	204	10	US-09-764-853-769	Sequence 769, App	763	5	2.6	227	10	US-09-846-573B-15	Sequence 15, Appl
691	5	2.6	205	9	US-10-025-380-1108	Sequence 1108, Ap	764	5	2.6	227	10	US-09-925-300-1124	Sequence 1124, Ap
692	5	2.6	205	10	US-09-922-217-1108	Sequence 1108, Ap	765	5	2.6	229	9	US-10-001-189-62	Sequence 62, Appl
693	5	2.6	206	9	US-09-764-868-679	Sequence 679, App	766	5	2.6	229	9	US-10-001-189-66	Sequence 66, Appl
694	5	2.6	209	9	US-09-738-626-4173	Sequence 4173, Ap	767	5	2.6	229	9	US-09-738-626-4779	Sequence 4779, Ap
695	5	2.6	210	10	US-09-840-787-6	Sequence 6, Appli	768	5	2.6	230	10	US-09-815-242-11783	Sequence 11783, A
696	5	2.6	210	10	US-09-733-167-6	Sequence 6, Appli	769	5	2.6	230	10	US-09-815-242-13533	Sequence 13533, A
697	5	2.6	210	10	US-09-881-752A-90	Sequence 90, Appl	770	5	2.6	232	9	US-09-792-251-8	Sequence 8, Appli
698	5	2.6	211	10	US-09-846-573B-12	Sequence 12, Appl	771	5	2.6	233	10	US-09-799-777-65	Sequence 65, Appl
699	5	2.6	212	9	US-09-738-626-5989	Sequence 5989, Ap	772	5	2.6	234	10	US-09-941-831-23	Sequence 23, Appl
700	5	2.6	213	10	US-09-815-242-10874	Sequence 10874, A	773	5	2.6	234	10	US-09-815-242-10297	Sequence 10297, A
701	5	2.6	213	12	US-10-001-843-178	Sequence 178, App	774	5	2.6	234	10	US-09-815-242-13737	Sequence 13737, A
702	5	2.6	214	9	US-10-091-613-1	Sequence 1, Appli	775	5	2.6	234	10	US-09-764-877-1775	Sequence 1775, Ap
703	5	2.6	214	10	US-09-198-559-1	Sequence 1, Appli	776	5	2.6	235	9	US-10-210-296-4	Sequence 4, Appli
704	5	2.6	214	10	US-09-747-155-66	Sequence 66, Appl	777	5	2.6	236	10	US-09-859-053-38	Sequence 38, Appl
705	5	2.6	215	9	US-09-992-598-389	Sequence 389, App	778	5	2.6	236	10	US-09-859-053-34	Sequence 34, Appl
706	5	2.6	215	9	US-09-989-293A-389	Sequence 389, App	779	5	2.6	237	9	US-10-135-965-1	Sequence 1, Appli
707	5	2.6	215	9	US-09-989-735-389	Sequence 389, App	780	5	2.6	237	9	US-10-135-965-46	Sequence 46, Appl
708	5	2.6	215	9	US-09-990-444-389	Sequence 389, App	781	5	2.6	238	9	US-09-854-133-37	Sequence 37, Appl
709	5	2.6	215	9	US-09-989-730-389	Sequence 389, App	782	5	2.6	238	10	US-09-738-973-37	Sequence 37, Appl
710	5	2.6	215	9	US-09-990-436-389	Sequence 389, App	783	5	2.6	239	9	US-09-738-626-5459	Sequence 5459, Ap
711	5	2.6	215	9	US-09-991-181-389	Sequence 389, App	784	5	2.6	239	10	US-09-925-637-54	Sequence 54, Appl
712	5	2.6	215	9	US-09-993-687-389	Sequence 389, App	785	5	2.6	240	10	US-09-922-683-8	Sequence 8, Appli
713	5	2.6	215	9	US-09-989-734-389	Sequence 389, App	786	5	2.6	240	10	US-09-920-552-61	Sequence 61, Appl
714	5	2.6	215	9	US-10-028-072-488	Sequence 488, App	787	5	2.6	241	9	US-09-813-453A-63	Sequence 63, Appl
715	5	2.6	215	9	US-09-997-653-389	Sequence 389, App	788	5	2.6	241	9	US-10-225-273-3	Sequence 3, Appli
716	5	2.6	215	9	US-09-993-667-389	Sequence 389, App	789	5	2.6	241	10	US-09-815-242-13948	Sequence 13948, A
717	5	2.6	215	9	US-10-121-049-488	Sequence 488, App	790	5	2.6	243	9	US-09-854-133-39	Sequence 39, Appl
718	5	2.6	215	9	US-10-123-904-488	Sequence 488, App	791	5	2.6	243	9	US-09-854-133-42	Sequence 42, Appl
719	5	2.6	215	9	US-10-140-470-488	Sequence 488, App	792	5	2.6	243	10	US-09-925-301-1055	Sequence 1055, Ap
720	5	2.6	215	9	US-09-990-438-389	Sequence 389, App	793	5	2.6	243	10	US-09-738-973-39	Sequence 39, Appl
721	5	2.6	215	9	US-09-990-562-389	Sequence 389, App	794	5	2.6	243	10	US-09-738-973-42	Sequence 42, Appl
722	5	2.6	215	9	US-09-997-428-389	Sequence 389, App	795	5	2.6	244	9	US-09-854-133-43	Sequence 43, Appl
723	5	2.6	215	9	US-09-997-666-389	Sequence 389, App	796	5	2.6	244	9	US-09-854-133-46	Sequence 46, Appl
724	5	2.6	215	9	US-10-176-746-488	Sequence 488, App	797	5	2.6	244	9	US-09-893-519A-17	Sequence 17, Appl
725	5	2.6	215	9	US-10-176-918-488	Sequence 488, App	798	5	2.6	244	10	US-09-738-973-43	Sequence 43, Appl
726	5	2.6	215	9	US-10-176-921-488	Sequence 488, App	799	5	2.6	244	10	US-09-738-973-46	Sequence 46, Appl
727	5	2.6	215	9	US-10-227-884-168	Sequence 168, App	800	5	2.6	245	9	US-09-995-913A-48	Sequence 48, Appl
728	5	2.6	215	9	US-09-990-711-389	Sequence 389, App	801	5	2.6	245	9	US-09-991-470-27	Sequence 27, Appl
729	5	2.6	215	9	US-10-137-865-488	Sequence 488, App	802	5	2.6	245	9	US-09-854-133-40	Sequence 40, Appl
730	5	2.6	215	9	US-10-140-474-488	Sequence 488, App	803	5	2.6	245	10	US-09-925-302-686	Sequence 686, App
731	5	2.6	215	9	US-10-142-431-488	Sequence 488, App	804	5	2.6	245	10	US-09-738-973-40	Sequence 40, Appl
732	5	2.6	215	9	US-10-143-114-488	Sequence 488, App	805	5	2.6	246	9	US-09-738-626-5141	Sequence 5141, Ap
733	5	2.6	215	9	US-10-230-163-168	Sequence 168, App	806	5	2.6	246	10	US-09-940-836A-6	Sequence 6, Appli
734	5	2.6	215	10	US-09-050-516-47	Sequence 47, Appl	807	5	2.6	247	9	US-10-135-965-3	Sequence 3, Appli
735	5	2.6	215	10	US-09-989-723-389	Sequence 389, App	808	5	2.6	247	10	US-09-925-302-517	Sequence 517, App
736	5	2.6	215	10	US-09-989-723-389	Sequence 389, App	809	5	2.6	248	9	US-10-036-041-59	Sequence 59, Appl
737	5	2.6	215	10	US-09-989-279-389	Sequence 389, App	810	5	2.6	248	9	US-10-035-855-59	Sequence 59, Appl
738	5	2.6	215	10	US-09-989-721-389	Sequence 389, App	811	5	2.6	248	9	US-10-174-590-480	Sequence 480, App
739	5	2.6	215	10	US-09-989-731-389	Sequence 389, App	812	5	2.6	248	9	US-10-176-758-480	Sequence 480, App
740	5	2.6	215	10	US-09-764-587A-4	Sequence 4, Appli	813	5	2.6	248	9	US-10-175-737-480	Sequence 480, App
741	5	2.6	215	10	US-09-989-733-389	Sequence 389, App	814	5	2.6	248	9	US-10-173-706-480	Sequence 480, App
742	5	2.6	215	10	US-09-991-073-389	Sequence 389, App	815	5	2.6	248	9	US-10-175-738-480	Sequence 480, App
743	5	2.6	215	10	US-09-990-442-389	Sequence 389, App	816	5	2.6	248	9	US-10-175-752-480	Sequence 480, App
744	5	2.6	215	10	US-09-991-163-389	Sequence 389, App	817	5	2.6	248	9	US-10-176-482-480	Sequence 480, App
745	5	2.6	215	10	US-09-993-604-389	Sequence 389, App	818	5	2.6	248	9	US-10-176-757-480	Sequence 480, App
746	5	2.6	215	10	US-09-990-456-389	Sequence 389, App	819	5	2.6	248	9	US-10-176-913-480	Sequence 480, App
747	5	2.6	215	10	US-09-989-721-389	Sequence 389, App	820	5	2.6	248	9	US-10-180-552-480	Sequence 480, App
748	5	2.6	217	9	US-09-738-626-6726	Sequence 6726, Ap	821	5	2.6	248	9	US-10-180-557-480	Sequence 480, App
749	5	2.6	220	9	US-09-738-626-6772	Sequence 6772, Ap	822	5	2.6	248	9	US-09-931-836-59	Sequence 59, Appl

531	5	2.6	140	9	US-10-184-646-300	Sequence 300, App	604	5	2.6	146	10	US-09-932-679-2	Sequence 2, Appli
532	5	2.6	140	9	US-10-184-647-300	Sequence 300, App	605	5	2.6	147	10	US-09-864-761-39307	Sequence 39307, A
533	5	2.6	140	9	US-10-184-652-300	Sequence 300, App	606	5	2.6	147	10	US-09-815-242-5241	Sequence 5241, Ap
534	5	2.6	140	9	US-10-187-594-300	Sequence 300, App	607	5	2.6	147	10	US-09-815-242-12362	Sequence 12362, A
535	5	2.6	140	9	US-10-187-596-300	Sequence 300, App	608	5	2.6	147	10	US-09-778-927A-75	Sequence 75, Appl
536	5	2.6	140	9	US-10-187-745-300	Sequence 300, App	609	5	2.6	147	10	US-09-778-927A-75	Sequence 76, Appl
537	5	2.6	140	9	US-10-187-886-300	Sequence 300, App	610	5	2.6	148	9	US-09-738-626-3643	Sequence 3643, Ap
538	5	2.6	140	9	US-10-187-885-300	Sequence 300, App	611	5	2.6	148	10	US-09-970-711-18	Sequence 18, Appl
539	5	2.6	140	9	US-10-195-464-300	Sequence 300, App	612	5	2.6	149	10	US-09-934-465-17	Sequence 17, Appl
540	5	2.6	140	9	US-10-176-751-300	Sequence 300, App	613	5	2.6	150	9	US-09-738-626-3936	Sequence 3936, Ap
541	5	2.6	140	9	US-10-176-760-300	Sequence 300, App	614	5	2.6	150	9	US-09-782-397-5	Sequence 5, Appli
542	5	2.6	140	9	US-10-176-990-300	Sequence 300, App	615	5	2.6	150	10	US-09-864-761-43913	Sequence 43913, A
543	5	2.6	140	9	US-10-180-541-300	Sequence 300, App	616	5	2.6	150	12	US-10-001-879-170	Sequence 170, App
544	5	2.6	140	9	US-10-180-542-300	Sequence 300, App	617	5	2.6	151	12	US-10-001-879-188	Sequence 188, App
545	5	2.6	140	9	US-10-180-548-300	Sequence 300, App	618	5	2.6	151	12	US-10-001-843-213	Sequence 213, App
546	5	2.6	140	9	US-10-180-551-300	Sequence 300, App	619	5	2.6	152	10	US-09-764-877-1129	Sequence 1129, Ap
547	5	2.6	140	9	US-10-180-998-300	Sequence 300, App	620	5	2.6	153	9	US-09-796-692-1455	Sequence 1455, Ap
548	5	2.6	140	9	US-10-180-999-300	Sequence 300, App	621	5	2.6	153	9	US-09-796-692-2481	Sequence 2481, Ap
549	5	2.6	140	9	US-10-183-012-300	Sequence 300, App	622	5	2.6	155	9	US-10-001-835-213	Sequence 213, App
550	5	2.6	140	9	US-10-184-613-300	Sequence 300, App	623	5	2.6	155	9	US-09-860-670-87	Sequence 87, Appl
551	5	2.6	140	9	US-10-184-616-300	Sequence 300, App	624	5	2.6	155	10	US-09-867-550-1948	Sequence 1948, Ap
552	5	2.6	140	9	US-10-184-617-300	Sequence 300, App	625	5	2.6	156	9	US-10-053-200-2	Sequence 2, Appli
553	5	2.6	140	9	US-10-184-622-300	Sequence 300, App	626	5	2.6	156	10	US-09-815-242-11752	Sequence 11752, A
554	5	2.6	140	9	US-10-184-628-300	Sequence 300, App	627	5	2.6	157	9	US-09-738-626-6808	Sequence 6808, Ap
555	5	2.6	140	9	US-10-184-629-300	Sequence 300, App	628	5	2.6	157	10	US-09-741-669-1366	Sequence 366, App
556	5	2.6	140	9	US-10-184-630-300	Sequence 300, App	629	5	2.6	157	10	US-09-912-020-374	Sequence 374, App
557	5	2.6	140	9	US-10-184-631-300	Sequence 300, App	630	5	2.6	158	9	US-10-260-877-30	Sequence 30, Appl
558	5	2.6	140	9	US-10-184-632-300	Sequence 300, App	631	5	2.6	160	9	US-09-854-133-205	Sequence 205, App
559	5	2.6	140	9	US-10-184-636-300	Sequence 300, App	632	5	2.6	160	10	US-09-815-242-11231	Sequence 11231, A
560	5	2.6	140	9	US-10-184-640-300	Sequence 300, App	633	5	2.6	160	10	US-09-916-790-35	Sequence 35, Appl
561	5	2.6	140	9	US-10-184-650-300	Sequence 300, App	634	5	2.6	160	10	US-09-738-973-205	Sequence 205, App
562	5	2.6	140	9	US-10-184-651-300	Sequence 300, App	635	5	2.6	161	10	US-09-965-602-22	Sequence 22, Appl
563	5	2.6	140	9	US-10-187-588-300	Sequence 300, App	636	5	2.6	163	9	US-10-125-258-82	Sequence 82, Appl
564	5	2.6	140	9	US-10-187-597-300	Sequence 300, App	637	5	2.6	164	10	US-09-863-824-4	Sequence 4, Appli
565	5	2.6	140	9	US-10-187-598-300	Sequence 300, App	638	5	2.6	165	9	US-10-125-258-83	Sequence 83, Appl
566	5	2.6	140	9	US-10-187-600-300	Sequence 300, App	639	5	2.6	166	9	US-09-899-046-152	Sequence 152, App
567	5	2.6	140	9	US-10-187-601-300	Sequence 300, App	640	5	2.6	166	9	US-09-878-281-152	Sequence 152, App
568	5	2.6	140	9	US-10-187-602-300	Sequence 300, App	641	5	2.6	167	10	US-09-764-846-143	Sequence 143, App
569	5	2.6	140	9	US-10-187-603-300	Sequence 300, App	642	5	2.6	168	9	US-09-764-868-687	Sequence 687, App
570	5	2.6	140	9	US-10-187-741-300	Sequence 300, App	643	5	2.6	168	9	US-09-798-889-51	Sequence 51, Appl
571	5	2.6	140	9	US-10-187-743-300	Sequence 300, App	644	5	2.6	169	9	US-09-899-046-42	Sequence 42, Appl
572	5	2.6	140	9	US-10-187-746-300	Sequence 300, App	645	5	2.6	169	9	US-09-899-046-44	Sequence 44, Appl
573	5	2.6	140	9	US-10-187-747-300	Sequence 300, App	646	5	2.6	169	9	US-09-878-281-42	Sequence 42, Appl
574	5	2.6	140	9	US-10-187-751-300	Sequence 300, App	647	5	2.6	169	9	US-09-878-281-47	Sequence 47, Appl
575	5	2.6	140	9	US-10-187-753-300	Sequence 300, App	648	5	2.6	169	10	US-09-938-803-2	Sequence 2, Appli
576	5	2.6	140	9	US-10-187-754-300	Sequence 300, App	649	5	2.6	171	9	US-09-738-626-6387	Sequence 6387, Ap
577	5	2.6	140	9	US-10-187-757-300	Sequence 300, App	650	5	2.6	172	10	US-09-800-729-138	Sequence 138, App
578	5	2.6	140	9	US-10-187-884-300	Sequence 300, App	651	5	2.6	172	10	US-09-764-847-694	Sequence 694, App
579	5	2.6	140	9	US-10-188-767-300	Sequence 300, App	652	5	2.6	173	10	US-09-764-877-1717	Sequence 1717, Ap
580	5	2.6	140	9	US-10-188-769-300	Sequence 300, App	653	5	2.6	174	9	US-09-764-868-633	Sequence 633, App
581	5	2.6	140	9	US-10-188-770-300	Sequence 300, App	654	5	2.6	174	10	US-09-731-872-264	Sequence 264, App
582	5	2.6	140	9	US-10-188-773-300	Sequence 300, App	655	5	2.6	176	10	US-09-764-869-1084	Sequence 1084, Ap
583	5	2.6	140	9	US-10-188-781-300	Sequence 300, App	656	5	2.6	177	10	US-09-764-864-942	Sequence 942, App
584	5	2.6	140	9	US-10-194-361-300	Sequence 300, App	657	5	2.6	178	10	US-09-811-284-256	Sequence 256, App
585	5	2.6	140	9	US-10-194-423-300	Sequence 300, App	658	5	2.6	178	10	US-09-764-853-469	Sequence 469, App
586	5	2.6	140	9	US-10-195-897-300	Sequence 300, App	659	5	2.6	179	9	US-09-738-626-5602	Sequence 5602, Ap
587	5	2.6	140	9	US-10-195-901-300	Sequence 300, App	660	5	2.6	179	10	US-09-815-242-4937	Sequence 4937, Ap
588	5	2.6	140	9	US-10-196-756-300	Sequence 300, App	661	5	2.6	179	10	US-09-815-242-10527	Sequence 10527, A
589	5	2.6	140	10	US-09-062-104A-2	Sequence 2, Appli	662	5	2.6	181	10	US-09-864-761-35721	Sequence 35721, A
590	5	2.6	140	12	US-10-006-867-90	Sequence 90, Appl	663	5	2.6	181	12	US-10-096-241-4	Sequence 4, Appli
591	5	2.6	140	12	US-10-052-586-300	Sequence 300, App	664	5	2.6	183	10	US-09-731-872-420	Sequence 420, App
592	5	2.6	140	12	US-10-116-378-30	Sequence 30, Appl	665	5	2.6	185	10	US-09-815-242-11083	Sequence 11083, A
593	5	2.6	141	10	US-09-877-156-22	Sequence 22, Appl	666	5	2.6	186	10	US-09-848-852A-5	Sequence 5, Appli
594	5	2.6	141	12	US-10-001-879-186	Sequence 186, App	667	5	2.6	188	10	US-09-925-299-1255	Sequence 1255, Ap
595	5	2.6	142	10	US-09-216-393-24	Sequence 24, Appl	668	5	2.6	192	9	US-09-832-355A-107	Sequence 107, App
596	5	2.6	143	8	US-08-981-087A-4	Sequence 4, Appli	669	5	2.6	192	9	US-09-832-355A-110	Sequence 110, App
597	5	2.6	143	9	US-09-779-050A-8	Sequence 8, Appli	670	5	2.6	192	10	US-09-772-105-18	Sequence 18, Appl
598	5	2.6	143	9	US-09-779-050A-9	Sequence 9, Appli	671	5	2.6	192	10	US-09-966-608-8	Sequence 8, Appli
599	5	2.6	144	8	US-08-981-087A-2	Sequence 2, Appli	672	5	2.6	193	9	US-09-895-913A-288	Sequence 288, App
600	5	2.6	145	10	US-09-728-479-8	Sequence 8, Appli	673	5	2.6	193	9	US-09-738-626-4529	Sequence 4529, Ap
601	5	2.6	145	10	US-09-894-526-5	Sequence 5, Appli	674	5	2.6	193	10	US-09-828-644-91	Sequence 91, Appl
602	5	2.6	145	10	US-09-263-689-12	Sequence 12, Appl	675	5	2.6	196	10	US-09-815-242-5413	Sequence 5413, Ap
603	5	2.6	146	9	US-10-260-877-60	Sequence 60, Appl	676	5	2.6	196	10	US-09-815-242-10825	Sequence 10825, A

385	5	2.6	96	10	US-09-815-242-10551	Sequence 10551, A	458	5	2.6	131	10	US-09-925-302-466	Sequence 466, App
386	5	2.6	98	9	US-09-991-496-42	Sequence 42, App	459	5	2.6	132	10	US-09-815-242-5913	Sequence 5913, Ap
387	5	2.6	98	9	US-10-001-883-133	Sequence 133, App	460	5	2.6	132	10	US-09-815-242-13155	Sequence 13155, A
388	5	2.6	98	10	US-09-864-761-37789	Sequence 37789, A	461	5	2.6	133	9	US-09-738-626-6531	Sequence 6531, Ap
389	5	2.6	98	10	US-09-905-243-45	Sequence 45, Appl	462	5	2.6	134	10	US-09-800-729-141	Sequence 141, App
390	5	2.6	98	10	US-09-874-923-42	Sequence 42, Appl	463	5	2.6	135	9	US-10-013-379-35	Sequence 35, Appl
391	5	2.6	98	10	US-09-925-300-1727	Sequence 1727, Ap	464	5	2.6	135	9	US-09-832-355A-111	Sequence 111, App
392	5	2.6	100	10	US-09-864-761-41304	Sequence 41304, A	465	5	2.6	137	10	US-09-864-761-33992	Sequence 33992, A
393	5	2.6	100	10	US-09-899-896-4	Sequence 4, Appl	466	5	2.6	139	10	US-09-939-980-409	Sequence 409, App
394	5	2.6	101	9	US-09-738-626-4069	Sequence 4069, Ap	467	5	2.6	139	12	US-10-096-241-33	Sequence 33, Appl
395	5	2.6	102	9	US-09-738-626-4790	Sequence 4790, Ap	468	5	2.6	140	9	US-10-063-547-90	Sequence 90, Appl
396	5	2.6	102	10	US-09-864-761-48557	Sequence 48557, A	469	5	2.6	140	9	US-10-174-590-300	Sequence 300, App
397	5	2.6	102	10	US-09-867-550-674	Sequence 674, App	470	5	2.6	140	9	US-10-176-758-300	Sequence 300, App
398	5	2.6	104	10	US-09-867-550-310	Sequence 310, App	471	5	2.6	140	9	US-10-063-616-90	Sequence 90, Appl
399	5	2.6	104	10	US-09-828-708-6	Sequence 6, Appl	472	5	2.6	140	9	US-10-175-737-300	Sequence 300, App
400	5	2.6	105	10	US-09-205-658-72	Sequence 72, Appl	473	5	2.6	140	9	US-10-175-737-300	Sequence 300, App
401	5	2.6	105	10	US-09-844-353A-72	Sequence 72, Appl	474	5	2.6	140	9	US-10-063-503-90	Sequence 90, Appl
402	5	2.6	105	10	US-09-764-869-1175	Sequence 1175, Ap	475	5	2.6	140	9	US-10-173-706-300	Sequence 300, App
403	5	2.6	105	10	US-09-828-708-3	Sequence 3, Appl	476	5	2.6	140	9	US-10-175-738-300	Sequence 300, App
404	5	2.6	105	10	US-09-828-708-7	Sequence 7, Appl	477	5	2.6	140	9	US-10-175-732-300	Sequence 300, App
405	5	2.6	106	8	US-08-844-215-14	Sequence 14, Appl	478	5	2.6	140	9	US-10-176-482-300	Sequence 300, App
406	5	2.6	106	9	US-10-013-379-43	Sequence 43, Appl	479	5	2.6	140	9	US-10-176-757-300	Sequence 300, App
407	5	2.6	106	9	US-09-738-626-4954	Sequence 4954, Ap	480	5	2.6	140	9	US-10-176-913-300	Sequence 300, App
408	5	2.6	106	10	US-09-764-877-1980	Sequence 1980, Ap	481	5	2.6	140	9	US-10-180-553-300	Sequence 300, App
409	5	2.6	107	8	US-08-844-215-8	Sequence 8, Appl	482	5	2.6	140	9	US-10-180-557-300	Sequence 300, App
410	5	2.6	107	8	US-08-844-215-10	Sequence 10, Appl	483	5	2.6	140	9	US-10-173-700-300	Sequence 300, App
411	5	2.6	107	8	US-08-844-215-12	Sequence 12, Appl	484	5	2.6	140	9	US-10-174-572-300	Sequence 300, App
412	5	2.6	107	9	US-09-991-470-25	Sequence 25, Appl	485	5	2.6	140	9	US-10-174-579-300	Sequence 300, App
413	5	2.6	107	9	US-09-796-692-1468	Sequence 1468, Ap	486	5	2.6	140	9	US-10-174-583-300	Sequence 300, App
414	5	2.6	107	9	US-09-796-692-1910	Sequence 1910, Ap	487	5	2.6	140	9	US-10-174-588-300	Sequence 300, App
415	5	2.6	107	9	US-09-796-692-1950	Sequence 1950, Ap	488	5	2.6	140	9	US-10-175-739-300	Sequence 300, App
416	5	2.6	107	10	US-09-864-761-45996	Sequence 45996, A	489	5	2.6	140	9	US-10-175-743-300	Sequence 300, App
417	5	2.6	108	9	US-09-974-879-149	Sequence 149, App	490	5	2.6	140	9	US-10-175-743-300	Sequence 300, App
418	5	2.6	108	10	US-09-134-333-10	Sequence 10, Appl	491	5	2.6	140	9	US-10-176-488-300	Sequence 300, App
419	5	2.6	108	10	US-09-764-864-1388	Sequence 1388, Ap	492	5	2.6	140	9	US-10-176-492-300	Sequence 300, App
420	5	2.6	110	10	US-09-864-761-42215	Sequence 42215, A	493	5	2.6	140	9	US-10-176-747-300	Sequence 300, App
421	5	2.6	110	10	US-09-932-679-52	Sequence 52, Appl	494	5	2.6	140	9	US-10-176-750-300	Sequence 300, App
422	5	2.6	110	10	US-09-939-980-391	Sequence 391, App	495	5	2.6	140	9	US-10-176-985-300	Sequence 300, App
423	5	2.6	111	10	US-09-864-761-36932	Sequence 36932, A	496	5	2.6	140	9	US-10-176-991-300	Sequence 300, App
424	5	2.6	111	10	US-09-864-761-40073	Sequence 40073, A	497	5	2.6	140	9	US-10-176-992-300	Sequence 300, App
425	5	2.6	111	10	US-09-864-761-44103	Sequence 44103, A	498	5	2.6	140	9	US-10-176-993-300	Sequence 300, App
426	5	2.6	111	9	US-09-867-550-1868	Sequence 1868, Ap	499	5	2.6	140	9	US-10-176-993-300	Sequence 300, App
427	5	2.6	114	9	US-09-764-868-1108	Sequence 1108, Ap	500	5	2.6	140	9	US-10-184-658-300	Sequence 300, App
428	5	2.6	114	10	US-09-263-959-305	Sequence 305, App	501	5	2.6	140	9	US-10-173-695-300	Sequence 300, App
429	5	2.6	114	10	US-09-263-959-306	Sequence 306, App	502	5	2.6	140	9	US-10-173-697-300	Sequence 300, App
430	5	2.6	114	10	US-09-263-959-307	Sequence 307, App	503	5	2.6	140	9	US-10-173-705-300	Sequence 300, App
431	5	2.6	114	10	US-09-263-959-308	Sequence 308, App	504	5	2.6	140	9	US-10-174-576-300	Sequence 300, App
432	5	2.6	114	10	US-09-263-959-309	Sequence 309, App	505	5	2.6	140	9	US-10-174-585-300	Sequence 300, App
433	5	2.6	115	10	US-09-764-860-394	Sequence 394, App	506	5	2.6	140	9	US-10-174-586-300	Sequence 300, App
434	5	2.6	116	10	US-09-864-761-44770	Sequence 44770, A	507	5	2.6	140	9	US-10-175-747-300	Sequence 300, App
435	5	2.6	116	10	US-09-778-927A-72	Sequence 72, Appl	508	5	2.6	140	9	US-10-176-481-300	Sequence 300, App
436	5	2.6	116	10	US-09-778-927A-73	Sequence 73, Appl	509	5	2.6	140	9	US-10-176-483-300	Sequence 300, App
437	5	2.6	116	10	US-09-134-333-5	Sequence 5, Appl	510	5	2.6	140	9	US-10-176-487-300	Sequence 300, App
438	5	2.6	117	10	US-09-810-560-9	Sequence 9, Appl	511	5	2.6	140	9	US-10-176-493-300	Sequence 300, App
439	5	2.6	117	10	US-09-815-242-12433	Sequence 12433, A	512	5	2.6	140	9	US-10-176-756-300	Sequence 300, App
440	5	2.6	117	10	US-09-815-242-12867	Sequence 12867, A	513	5	2.6	140	9	US-10-176-911-300	Sequence 300, App
441	5	2.6	117	10	US-09-815-242-13094	Sequence 13094, A	514	5	2.6	140	9	US-10-176-919-300	Sequence 300, App
442	5	2.6	118	9	US-09-738-626-6243	Sequence 6243, Ap	515	5	2.6	140	9	US-10-176-925-300	Sequence 300, App
443	5	2.6	119	10	US-09-815-242-5341	Sequence 5341, Ap	516	5	2.6	140	9	US-10-176-978-300	Sequence 300, App
444	5	2.6	119	10	US-09-815-242-12401	Sequence 12401, A	517	5	2.6	140	9	US-10-179-510-300	Sequence 300, App
445	5	2.6	121	10	US-09-771-383-5	Sequence 5, Appl	518	5	2.6	140	9	US-10-180-543-300	Sequence 300, App
446	5	2.6	121	10	US-09-864-761-45409	Sequence 45409, A	519	5	2.6	140	9	US-10-180-544-300	Sequence 300, App
447	5	2.6	123	10	US-09-757-049A-10	Sequence 10, Appl	520	5	2.6	140	9	US-10-180-546-300	Sequence 300, App
448	5	2.6	123	10	US-09-924-946-4	Sequence 4, Appl	521	5	2.6	140	9	US-10-180-547-300	Sequence 300, App
449	5	2.6	126	9	US-10-023-437-21	Sequence 21, Appl	522	5	2.6	140	9	US-10-180-549-300	Sequence 300, App
450	5	2.6	126	10	US-09-932-679-45	Sequence 45, Appl	523	5	2.6	140	9	US-10-180-555-300	Sequence 300, App
451	5	2.6	127	10	US-09-134-333-8	Sequence 8, Appl	524	5	2.6	140	9	US-10-180-559-300	Sequence 300, App
452	5	2.6	127	10	US-09-764-903-41	Sequence 41, Appl	525	5	2.6	140	9	US-10-181-000-300	Sequence 300, App
453	5	2.6	128	9	US-09-974-879-523	Sequence 523, App	526	5	2.6	140	9	US-10-183-010-300	Sequence 300, App
454	5	2.6	128	10	US-09-864-761-34009	Sequence 34009, A	527	5	2.6	140	9	US-10-183-012-300	Sequence 300, App
455	5	2.6	128	10	US-09-925-301-1373	Sequence 1373, Ap	528	5	2.6	140	9	US-10-184-614-300	Sequence 300, App
456	5	2.6	130	10	US-09-867-550-616	Sequence 616, App	529	5	2.6	140	9	US-10-184-623-300	Sequence 300, App
457	5	2.6	130	10	US-09-925-300-1796	Sequence 1796, Ap	530	5	2.6	140	9	US-10-184-635-300	Sequence 300, App

239	5	2.6	21	10	US-09-853-830-62	Sequence 62, Appl	312	5	2.6	58	9	US-09-974-879-168	Sequence 168, App
240	5	2.6	21	10	US-09-853-830-63	Sequence 63, Appl	313	5	2.6	58	10	US-09-864-761-35621	Sequence 35621, A
241	5	2.6	21	10	US-09-853-830-66	Sequence 66, Appl	314	5	2.6	58	10	US-09-864-761-47446	Sequence 47446, A
242	5	2.6	21	10	US-09-853-830-183	Sequence 183, App	315	5	2.6	61	9	US-10-001-857-138	Sequence 138, App
243	5	2.6	22	10	US-09-949-196-33	Sequence 33, Appl	316	5	2.6	61	10	US-09-841-132-6	Sequence 6, Appl
244	5	2.6	22	10	US-09-949-196-37	Sequence 37, Appl	317	5	2.6	62	10	US-09-864-761-40006	Sequence 40006, A
245	5	2.6	22	10	US-09-949-196-41	Sequence 41, Appl	318	5	2.6	62	10	US-09-864-761-48278	Sequence 48278, A
246	5	2.6	24	10	US-09-864-761-43949	Sequence 43949, A	319	5	2.6	62	10	US-09-867-550-1126	Sequence 1126, Ap
247	5	2.6	24	10	US-09-999-256-5	Sequence 5, Appl	320	5	2.6	63	9	US-09-796-692-1818	Sequence 1818, Ap
248	5	2.6	24	10	US-09-999-256-13	Sequence 13, Appl	321	5	2.6	63	9	US-09-796-692-2072	Sequence 2072, Ap
249	5	2.6	24	10	US-09-999-256-14	Sequence 14, Appl	322	5	2.6	63	9	US-09-796-692-2312	Sequence 2312, Ap
250	5	2.6	25	8	US-08-424-550B-288	Sequence 288, App	323	5	2.6	63	10	US-09-864-761-41188	Sequence 41188, A
251	5	2.6	25	10	US-09-798-831-1	Sequence 1, Appl	324	5	2.6	63	10	US-09-867-550-1798	Sequence 1798, Ap
252	5	2.6	25	10	US-09-798-831-5	Sequence 5, Appl	325	5	2.6	63	12	US-10-035-408-1	Sequence 1, Appl
253	5	2.6	25	10	US-09-798-831-6	Sequence 6, Appl	326	5	2.6	64	10	US-09-864-761-46964	Sequence 46964, A
254	5	2.6	25	10	US-09-798-831-7	Sequence 7, Appl	327	5	2.6	65	10	US-09-764-847-735	Sequence 735, App
255	5	2.6	25	10	US-09-999-256-16	Sequence 16, Appl	328	5	2.6	65	10	US-09-993-844-53	Sequence 53, Appl
256	5	2.6	28	8	US-08-908-864-6	Sequence 6, Appl	329	5	2.6	68	9	US-09-756-854-22	Sequence 22, Appl
257	5	2.6	28	10	US-09-908-323-6	Sequence 6, Appl	330	5	2.6	68	9	US-10-041-574-22	Sequence 22, Appl
258	5	2.6	29	9	US-09-974-879-527	Sequence 527, App	331	5	2.6	68	9	US-09-796-692-687	Sequence 687, App
259	5	2.6	29	10	US-09-864-761-38513	Sequence 38513, A	332	5	2.6	68	9	US-10-023-896-96	Sequence 96, Appl
260	5	2.6	30	10	US-09-864-761-38388	Sequence 38388, A	333	5	2.6	70	9	US-09-764-868-662	Sequence 662, App
261	5	2.6	31	10	US-09-050-516-44	Sequence 44, Appl	334	5	2.6	70	10	US-09-864-761-40856	Sequence 40856, A
262	5	2.6	32	10	US-09-864-761-43462	Sequence 43462, A	335	5	2.6	71	10	US-09-864-761-36820	Sequence 36820, A
263	5	2.6	32	10	US-09-828-708-97	Sequence 97, Appl	336	5	2.6	71	10	US-09-864-761-926	Sequence 926, App
264	5	2.6	32	10	US-09-828-708-100	Sequence 100, App	337	5	2.6	72	9	US-10-001-835-209	Sequence 209, App
265	5	2.6	32	10	US-09-828-708-101	Sequence 101, App	338	5	2.6	72	9	US-09-984-245-317	Sequence 317, App
266	5	2.6	33	8	US-08-908-884-8	Sequence 8, Appl	339	5	2.6	74	9	US-09-738-626-6030	Sequence 6030, Ap
267	5	2.6	33	10	US-09-864-761-44741	Sequence 44741, A	340	5	2.6	74	10	US-09-864-761-43233	Sequence 43233, A
268	5	2.6	33	10	US-09-908-323-8	Sequence 8, Appl	341	5	2.6	74	10	US-09-932-679-51	Sequence 51, Appl
269	5	2.6	35	10	US-09-925-300-1747	Sequence 1747, Ap	342	5	2.6	74	12	US-10-025-687-23	Sequence 23, Appl
270	5	2.6	36	10	US-09-764-877-1477	Sequence 1477, Ap	343	5	2.6	75	9	US-09-884-456-35	Sequence 35, Appl
271	5	2.6	37	9	US-09-738-626-6826	Sequence 6826, Ap	344	5	2.6	76	9	US-10-058-820-21	Sequence 21, Appl
272	5	2.6	38	9	US-09-259-658-8	Sequence 8, Appl	345	5	2.6	76	9	US-09-796-692-1076	Sequence 1076, Ap
273	5	2.6	38	10	US-09-864-761-36287	Sequence 36287, A	346	5	2.6	76	10	US-09-804-866-16	Sequence 16, Appl
274	5	2.6	38	10	US-09-864-761-44070	Sequence 44070, A	347	5	2.6	76	10	US-09-826-312-13	Sequence 13, Appl
275	5	2.6	39	10	US-09-864-761-40353	Sequence 40353, A	348	5	2.6	76	10	US-09-993-844-49	Sequence 49, Appl
276	5	2.6	40	10	US-09-864-761-40732	Sequence 40732, A	349	5	2.6	77	9	US-09-992-964-17	Sequence 17, Appl
277	5	2.6	41	10	US-09-864-761-47152	Sequence 47152, A	350	5	2.6	77	9	US-10-112-793-24	Sequence 24, Appl
278	5	2.6	42	10	US-09-864-761-47152	Sequence 47152, A	351	5	2.6	77	9	US-09-738-626-5825	Sequence 5825, Ap
279	5	2.6	44	10	US-09-864-761-45911	Sequence 45911, A	352	5	2.6	77	9	US-09-983-802-552	Sequence 552, App
280	5	2.6	44	10	US-09-864-761-36018	Sequence 36018, A	353	5	2.6	77	10	US-09-764-869-738	Sequence 738, App
281	5	2.6	44	10	US-09-864-761-36124	Sequence 36124, A	354	5	2.6	77	10	US-09-887-879-17	Sequence 17, Appl
282	5	2.6	45	10	US-09-864-761-38032	Sequence 38032, A	355	5	2.6	77	10	US-09-764-847-892	Sequence 892, App
283	5	2.6	45	10	US-09-932-679-50	Sequence 50, Appl	356	5	2.6	78	10	US-09-864-761-47107	Sequence 47107, A
284	5	2.6	46	10	US-09-864-761-44801	Sequence 44801, A	357	5	2.6	78	10	US-09-864-761-48441	Sequence 48441, A
285	5	2.6	47	10	US-09-864-761-37523	Sequence 37523, A	358	5	2.6	79	10	US-09-764-887-235	Sequence 235, App
286	5	2.6	47	10	US-09-864-761-46971	Sequence 46971, A	359	5	2.6	79	10	US-09-784-810A-10	Sequence 10, Appl
287	5	2.6	47	10	US-09-999-256-7	Sequence 7, Appl	360	5	2.6	79	12	US-10-001-843-148	Sequence 148, App
288	5	2.6	47	10	US-09-999-256-8	Sequence 8, Appl	361	5	2.6	81	10	US-09-925-300-1772	Sequence 1772, Ap
289	5	2.6	47	10	US-09-999-256-9	Sequence 9, Appl	362	5	2.6	81	10	US-09-858-664A-32	Sequence 32, Appl
290	5	2.6	48	10	US-09-864-761-35521	Sequence 35521, A	363	5	2.6	83	9	US-09-764-868-856	Sequence 856, App
291	5	2.6	49	10	US-09-864-761-48049	Sequence 48049, A	364	5	2.6	83	9	US-09-738-626-5034	Sequence 5034, Ap
292	5	2.6	50	9	US-09-974-879-257	Sequence 257, App	365	5	2.6	84	10	US-09-826-312-14	Sequence 14, Appl
293	5	2.6	50	9	US-09-995-419A-17	Sequence 17, Appl	366	5	2.6	84	10	US-09-864-761-41647	Sequence 41647, A
294	5	2.6	50	10	US-09-864-761-34382	Sequence 34382, A	367	5	2.6	84	10	US-09-925-300-1647	Sequence 1647, Ap
295	5	2.6	50	10	US-09-864-761-38314	Sequence 38314, A	368	5	2.6	84	12	US-10-105-992-6	Sequence 6, Appl
296	5	2.6	50	10	US-09-864-761-42116	Sequence 42116, A	369	5	2.6	85	10	US-09-826-312-15	Sequence 15, Appl
297	5	2.6	51	9	US-10-023-896-77	Sequence 77, Appl	370	5	2.6	85	10	US-09-864-761-40047	Sequence 40047, A
298	5	2.6	51	9	US-10-023-896-95	Sequence 95, Appl	371	5	2.6	85	10	US-09-867-550-1726	Sequence 1726, Ap
299	5	2.6	51	10	US-09-864-761-41212	Sequence 41212, A	372	5	2.6	86	10	US-09-841-132-5	Sequence 5, Appl
300	5	2.6	51	12	US-10-067-422-18	Sequence 18, Appl	373	5	2.6	87	9	US-10-125-258-73	Sequence 73, Appl
301	5	2.6	52	10	US-09-864-761-46780	Sequence 46780, A	374	5	2.6	87	9	US-10-125-258-74	Sequence 74, Appl
302	5	2.6	53	10	US-09-864-761-43592	Sequence 43592, A	375	5	2.6	88	10	US-09-864-761-42833	Sequence 42833, A
303	5	2.6	53	10	US-09-800-729-164	Sequence 164, App	376	5	2.6	88	10	US-09-764-853-591	Sequence 591, App
304	5	2.6	53	12	US-10-001-879-162	Sequence 162, App	377	5	2.6	89	10	US-09-815-242-4918	Sequence 4918, App
305	5	2.6	54	10	US-09-864-761-46487	Sequence 46487, A	378	5	2.6	90	9	US-09-854-286-14	Sequence 14, Appl
306	5	2.6	54	10	US-09-925-300-972	Sequence 972, App	379	5	2.6	90	9	US-10-002-974-26	Sequence 26, Appl
307	5	2.6	54	12	US-10-001-870-155	Sequence 155, App	380	5	2.6	90	10	US-09-864-761-46509	Sequence 46509, A
308	5	2.6	56	10	US-09-864-761-42610	Sequence 42610, A	381	5	2.6	90	12	US-10-014-269-26	Sequence 26, Appl
309	5	2.6	56	10	US-09-864-761-42704	Sequence 42704, A	382	5	2.6	94	10	US-09-864-761-40525	Sequence 40525, A
310	5	2.6	57	10	US-09-864-761-35907	Sequence 35907, A	383	5	2.6	94	10	US-09-134-333-2	Sequence 2, Appl
311	5	2.6	57	10	US-09-864-761-42965	Sequence 42965, A	384	5	2.6	96	9	US-09-978-756-1	Sequence 1, Appl

93	6	3.1	472	9	US-10-184-658-98	Sequence 98, Appl	166	6	3.1	472	9	US-10-187-746-98	Sequence 98, Appl
94	6	3.1	472	9	US-10-173-695-98	Sequence 98, Appl	167	6	3.1	472	9	US-10-187-747-98	Sequence 98, Appl
95	6	3.1	472	9	US-10-173-697-98	Sequence 98, Appl	168	6	3.1	472	9	US-10-187-751-98	Sequence 98, Appl
96	6	3.1	472	9	US-10-173-705-98	Sequence 98, Appl	169	6	3.1	472	9	US-10-187-753-98	Sequence 98, Appl
97	6	3.1	472	9	US-10-174-576-98	Sequence 98, Appl	170	6	3.1	472	9	US-10-187-754-98	Sequence 98, Appl
98	6	3.1	472	9	US-10-174-585-98	Sequence 98, Appl	171	6	3.1	472	9	US-10-187-757-98	Sequence 98, Appl
99	6	3.1	472	9	US-10-174-586-98	Sequence 98, Appl	172	6	3.1	472	9	US-10-187-757-98	Sequence 98, Appl
100	6	3.1	472	9	US-10-175-747-98	Sequence 98, Appl	173	6	3.1	472	9	US-10-188-767-98	Sequence 98, Appl
101	6	3.1	472	9	US-10-176-481-98	Sequence 98, Appl	174	6	3.1	472	9	US-10-188-769-98	Sequence 98, Appl
102	6	3.1	472	9	US-10-176-485-98	Sequence 98, Appl	175	6	3.1	472	9	US-10-188-770-98	Sequence 98, Appl
103	6	3.1	472	9	US-10-176-487-98	Sequence 98, Appl	176	6	3.1	472	9	US-10-188-773-98	Sequence 98, Appl
104	6	3.1	472	9	US-10-176-493-98	Sequence 98, Appl	177	6	3.1	472	9	US-10-188-781-98	Sequence 98, Appl
105	6	3.1	472	9	US-10-176-756-98	Sequence 98, Appl	178	6	3.1	472	9	US-10-194-361-98	Sequence 98, Appl
106	6	3.1	472	9	US-10-176-911-98	Sequence 98, Appl	179	6	3.1	472	9	US-10-194-423-98	Sequence 98, Appl
107	6	3.1	472	9	US-10-176-919-98	Sequence 98, Appl	180	6	3.1	472	9	US-10-195-897-98	Sequence 98, Appl
108	6	3.1	472	9	US-10-176-925-98	Sequence 98, Appl	181	6	3.1	472	9	US-10-195-901-98	Sequence 98, Appl
109	6	3.1	472	9	US-10-176-978-98	Sequence 98, Appl	182	6	3.1	472	9	US-10-196-756-98	Sequence 98, Appl
110	6	3.1	472	9	US-10-179-510-98	Sequence 98, Appl	183	6	3.1	472	12	US-10-052-586-98	Sequence 98, Appl
111	6	3.1	472	9	US-10-180-543-98	Sequence 98, Appl	184	6	3.1	485	10	US-09-801-275-2	Sequence 2, Appl
112	6	3.1	472	9	US-10-180-544-98	Sequence 98, Appl	185	6	3.1	501	10	US-09-815-242-5292	Sequence 5292, Ap
113	6	3.1	472	9	US-10-180-546-98	Sequence 98, Appl	186	6	3.1	504	10	US-09-815-242-12503	Sequence 12503, A
114	6	3.1	472	9	US-10-180-547-98	Sequence 98, Appl	187	6	3.1	520	9	US-10-157-855-2	Sequence 2, Appl
115	6	3.1	472	9	US-10-180-549-98	Sequence 98, Appl	188	6	3.1	568	10	US-09-945-676-7	Sequence 7, Appl
116	6	3.1	472	9	US-10-180-555-98	Sequence 98, Appl	189	6	3.1	600	9	US-09-893-519A-31	Sequence 31, Appl
117	6	3.1	472	9	US-10-180-559-98	Sequence 98, Appl	190	6	3.1	607	9	US-10-072-436-11	Sequence 11, Appl
118	6	3.1	472	9	US-10-181-000-98	Sequence 98, Appl	191	6	3.1	618	12	US-10-001-851-28	Sequence 28, Appl
119	6	3.1	472	9	US-10-183-010-98	Sequence 98, Appl	192	6	3.1	650	10	US-09-943-671-19	Sequence 29, Appl
120	6	3.1	472	9	US-10-183-012-98	Sequence 98, Appl	193	6	3.1	752	10	US-09-825-809-2	Sequence 2, Appl
121	6	3.1	472	9	US-10-184-614-98	Sequence 98, Appl	194	6	3.1	752	10	US-09-825-809-4	Sequence 4, Appl
122	6	3.1	472	9	US-10-184-623-98	Sequence 98, Appl	195	6	3.1	755	9	US-09-738-626-3935	Sequence 3935, Ap
123	6	3.1	472	9	US-10-184-635-98	Sequence 98, Appl	196	6	3.1	764	10	US-09-866-582-41	Sequence 41, Appl
124	6	3.1	472	9	US-10-184-637-98	Sequence 98, Appl	197	6	3.1	787	9	US-09-924-340-70	Sequence 70, Appl
125	6	3.1	472	9	US-10-184-646-98	Sequence 98, Appl	198	6	3.1	787	9	US-09-992-600A-70	Sequence 70, Appl
126	6	3.1	472	9	US-10-184-647-98	Sequence 98, Appl	199	6	3.1	787	9	US-10-227-884-90	Sequence 90, Appl
127	6	3.1	472	9	US-10-184-652-98	Sequence 98, Appl	200	6	3.1	787	9	US-10-230-163-90	Sequence 90, Appl
128	6	3.1	472	9	US-10-187-594-98	Sequence 98, Appl	201	6	3.1	827	10	US-09-815-242-11357	Sequence 11357, A
129	6	3.1	472	9	US-10-187-596-98	Sequence 98, Appl	202	6	3.1	868	10	US-09-391-340-2	Sequence 2, Appl
130	6	3.1	472	9	US-10-187-745-98	Sequence 98, Appl	203	6	3.1	868	10	US-09-948-369-2	Sequence 2, Appl
131	6	3.1	472	9	US-10-187-885-98	Sequence 98, Appl	204	6	3.1	906	9	US-10-060-230-19	Sequence 19, Appl
132	6	3.1	472	9	US-10-187-886-98	Sequence 98, Appl	205	6	3.1	906	9	US-10-060-230-20	Sequence 20, Appl
133	6	3.1	472	9	US-10-199-464-98	Sequence 98, Appl	206	6	3.1	906	9	US-10-060-230-21	Sequence 21, Appl
134	6	3.1	472	9	US-10-176-751-98	Sequence 98, Appl	207	6	3.1	906	10	US-10-060-230-22	Sequence 22, Appl
135	6	3.1	472	9	US-10-176-760-98	Sequence 98, Appl	208	6	3.1	908	10	US-09-376-045-6	Sequence 6, Appl
136	6	3.1	472	9	US-10-176-990-98	Sequence 98, Appl	209	6	3.1	921	9	US-09-738-626-5559	Sequence 5559, Ap
137	6	3.1	472	9	US-10-180-541-98	Sequence 98, Appl	210	6	3.1	950	10	US-09-815-242-11961	Sequence 11961, A
138	6	3.1	472	9	US-10-180-542-98	Sequence 98, Appl	211	6	3.1	1033	10	US-09-888-615-75	Sequence 75, Appl
139	6	3.1	472	9	US-10-180-548-98	Sequence 98, Appl	212	6	3.1	1054	10	US-09-376-045-4	Sequence 4, Appl
140	6	3.1	472	9	US-10-180-551-98	Sequence 98, Appl	213	6	3.1	1119	10	US-09-782-714-2	Sequence 2, Appl
141	6	3.1	472	9	US-10-180-998-98	Sequence 98, Appl	214	6	3.1	1564	10	US-09-801-368-244	Sequence 244, App
142	6	3.1	472	9	US-10-180-999-98	Sequence 98, Appl	215	6	3.1	3256	10	US-09-919-172-98	Sequence 98, Appl
143	6	3.1	472	9	US-10-183-013-98	Sequence 98, Appl	216	5	2.6	5	9	US-09-933-999A-21	Sequence 21, Appl
144	6	3.1	472	9	US-10-184-612-98	Sequence 98, Appl	217	5	2.6	7	9	US-10-040-572-1	Sequence 1, Appl
145	6	3.1	472	9	US-10-184-616-98	Sequence 98, Appl	218	5	2.6	11	10	US-09-734-520-95	Sequence 95, Appl
146	6	3.1	472	9	US-10-184-617-98	Sequence 98, Appl	219	5	2.6	11	12	US-10-012-034A-95	Sequence 95, Appl
147	6	3.1	472	9	US-10-184-622-98	Sequence 98, Appl	220	5	2.6	12	9	US-10-038-612-156	Sequence 156, App
148	6	3.1	472	9	US-10-184-628-98	Sequence 98, Appl	221	5	2.6	13	9	US-10-038-612-157	Sequence 157, App
149	6	3.1	472	9	US-10-184-629-98	Sequence 98, Appl	222	5	2.6	15	10	US-09-932-679-46	Sequence 46, Appl
150	6	3.1	472	9	US-10-184-630-98	Sequence 98, Appl	223	5	2.6	15	10	US-09-841-132-100	Sequence 100, App
151	6	3.1	472	9	US-10-184-631-98	Sequence 98, Appl	224	5	2.6	15	12	US-10-041-030-16	Sequence 16, Appl
152	6	3.1	472	9	US-10-184-632-98	Sequence 98, Appl	225	5	2.6	16	10	US-09-949-196-1	Sequence 1, Appl
153	6	3.1	472	9	US-10-184-636-98	Sequence 98, Appl	226	5	2.6	17	9	US-10-024-918-19	Sequence 19, Appl
154	6	3.1	472	9	US-10-184-640-98	Sequence 98, Appl	227	5	2.6	18	9	US-09-836-073-18	Sequence 18, Appl
155	6	3.1	472	9	US-10-184-650-98	Sequence 98, Appl	228	5	2.6	18	10	US-09-734-520-32	Sequence 32, Appl
156	6	3.1	472	9	US-10-184-651-98	Sequence 98, Appl	229	5	2.6	18	12	US-10-012-034A-32	Sequence 32, Appl
157	6	3.1	472	9	US-10-187-588-98	Sequence 98, Appl	230	5	2.6	19	8	US-08-424-550B-398	Sequence 398, App
158	6	3.1	472	9	US-10-187-597-98	Sequence 98, Appl	231	5	2.6	19	9	US-10-038-612-25	Sequence 25, Appl
159	6	3.1	472	9	US-10-187-598-98	Sequence 98, Appl	232	5	2.6	19	9	US-10-038-612-26	Sequence 26, Appl
160	6	3.1	472	9	US-10-187-600-98	Sequence 98, Appl	233	5	2.6	19	10	US-09-909-652-5	Sequence 5, Appl
161	6	3.1	472	9	US-10-187-601-98	Sequence 98, Appl	234	5	2.6	20	9	US-10-038-612-158	Sequence 158, App
162	6	3.1	472	9	US-10-187-602-98	Sequence 98, Appl	235	5	2.6	20	10	US-09-864-761-41462	Sequence 41462, A
163	6	3.1	472	9	US-10-187-603-98	Sequence 98, Appl	236	5	2.6	21	10	US-09-864-761-48960	Sequence 48960, A
164	6	3.1	472	9	US-10-187-741-98	Sequence 98, Appl	237	5	2.6	21	10	US-09-815-306-16	Sequence 16, Appl
165	6	3.1	472	9	US-10-187-743-98	Sequence 98, Appl	238	5	2.6	21	10	US-09-853-830-26	Sequence 26, Appl

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2003, 10:01:14 : Search time 14 seconds
(without alignments)
517.328 Million cell updates/sec

Title: US-09-868-352-23

Perfect score: 192

Sequence: 1 MRKEVTPMLNKNYKPGQF.....EODKEMTSKQHLFVRKN 192

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 174566 seqs, 37721826 residues

Word size : 0

Total number of hits satisfying chosen parameters: 174566

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Published_Applications_AA:*

1: /cgn2_6/ptodata/2/pubpaa/PCTUS_NEW_PUB.pep.*

2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*

3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*

4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*

5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*

6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*

7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*

8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*

9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*

10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*

11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*

12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*

13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*

14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	12	6.2	176	10 US-09-815-242-13414	Sequence 13414, A
2	12	6.2	176	10 US-09-815-242-13567	Sequence 13567, A
3	7	3.6	15	12 US-10-085-027-2	Sequence 2, Appl1
4	7	3.6	157	9 US-10-117-846-19	Sequence 19, Appl1
5	7	3.6	159	12 US-10-085-027-5	Sequence 5, Appl1
6	7	3.6	200	10 US-09-867-550-1934	Sequence 1934, Ap
7	7	3.6	317	9 US-10-146-474-6	Sequence 6, Appl1
8	7	3.6	782	10 US-09-866-582-16	Sequence 16, Appl1
9	7	3.6	1242	9 US-09-436-184-5	Sequence 5, Appl1
10	7	3.6	1242	10 US-09-903-248-5	Sequence 5, Appl1
11	7	3.6	1242	10 US-09-859-604-5	Sequence 5, Appl1
12	7	3.6	1242	10 US-09-903-063-5	Sequence 5, Appl1
13	7	3.6	1242	10 US-09-903-216-5	Sequence 5, Appl1
14	7	3.6	1242	10 US-09-903-199-5	Sequence 5, Appl1
15	7	3.6	1242	10 US-09-903-023-5	Sequence 5, Appl1
16	7	3.6	1242	12 US-10-085-027-1	Sequence 1, Appl1
17	6	3.1	19	9 US-10-038-612-27	Sequence 27, Appl1
18	6	3.1	25	10 US-09-071-838-217	Sequence 217, App
19	6	3.1	38	10 US-09-864-761-40249	Sequence 40249, A

6	3.1	42	10	US-09-864-761-37254	Sequence 37254, A
6	3.1	49	10	US-09-864-761-35132	Sequence 35132, A
6	3.1	73	10	US-09-764-878-190	Sequence 190, App
6	3.1	75	9	US-09-796-892-1527	Sequence 1527, Ap
6	3.1	88	10	US-09-764-869-945	Sequence 945, App
6	3.1	98	10	US-09-925-299-1067	Sequence 1067, Ap
6	3.1	138	10	US-09-864-761-33652	Sequence 33652, A
6	3.1	146	10	US-09-764-864-1566	Sequence 1566, Ap
6	3.1	156	10	US-09-925-300-1731	Sequence 1731, Ap
6	3.1	196	10	US-09-159-469-32	Sequence 32, Appl
6	3.1	196	10	US-09-159-469-58	Sequence 58, Appl
6	3.1	196	10	US-09-798-042-32	Sequence 32, Appl
6	3.1	196	10	US-09-798-042-58	Sequence 58, Appl
6	3.1	227	9	US-09-774-639-155	Sequence 155, App
6	3.1	235	9	US-09-738-826-4680	Sequence 4680, Ap
6	3.1	236	9	US-09-738-826-6304	Sequence 6304, Ap
6	3.1	238	12	US-10-024-579-16	Sequence 16, Appl
6	3.1	257	12	US-10-024-579-14	Sequence 14, Appl
6	3.1	258	10	US-09-752-639-150	Sequence 150, App
6	3.1	258	10	US-09-984-198-150	Sequence 150, App
6	3.1	264	12	US-10-024-579-12	Sequence 12, Appl
6	3.1	283	12	US-10-024-579-10	Sequence 10, Appl
6	3.1	299	10	US-09-761-288-30	Sequence 30, Appl
6	3.1	301	9	US-10-116-255-22	Sequence 22, Appl
6	3.1	306	10	US-09-925-297-601	Sequence 601, App
6	3.1	307	10	US-09-761-288-27	Sequence 27, Appl
6	3.1	313	10	US-09-810-264-30	Sequence 30, Appl
6	3.1	316	10	US-09-925-300-1355	Sequence 1355, Ap
6	3.1	330	10	US-09-801-368-258	Sequence 258, App
6	3.1	331	9	US-09-866-570A-59	Sequence 59, Appl
6	3.1	331	10	US-09-866-572A-59	Sequence 59, Appl
6	3.1	355	9	US-09-947-953-2	Sequence 2, Appl1
6	3.1	383	9	US-09-895-913A-152	Sequence 152, App
6	3.1	385	9	US-10-023-895A-2	Sequence 2, Appl1
6	3.1	399	9	US-09-764-868-1005	Sequence 1005, Ap
6	3.1	406	10	US-09-741-669-408	Sequence 408, App
6	3.1	413	10	US-09-821-803A-7	Sequence 7, Appl1
6	3.1	413	9	US-09-821-803A-8	Sequence 8, Appl1
6	3.1	415	9	US-09-764-868-1027	Sequence 1027, Ap
6	3.1	446	10	US-09-850-351A-52	Sequence 52, Appl
6	3.1	472	9	US-09-978-295A-245	Sequence 245, App
6	3.1	472	9	US-09-978-697-245	Sequence 245, App
6	3.1	472	9	US-09-978-192A-245	Sequence 245, App
6	3.1	472	9	US-09-999-832A-245	Sequence 245, App
6	3.1	472	9	US-09-978-189-245	Sequence 245, App
6	3.1	472	9	US-10-174-590-98	Sequence 98, Appl
6	3.1	472	9	US-10-176-758-98	Sequence 98, Appl
6	3.1	472	9	US-10-175-737-98	Sequence 98, Appl
6	3.1	472	9	US-10-173-706-98	Sequence 98, Appl
6	3.1	472	9	US-10-175-738-98	Sequence 98, Appl
6	3.1	472	9	US-10-175-752-98	Sequence 98, Appl
6	3.1	472	9	US-10-176-482-98	Sequence 98, Appl
6	3.1	472	9	US-10-176-757-98	Sequence 98, Appl
6	3.1	472	9	US-10-176-913-98	Sequence 98, Appl
6	3.1	472	9	US-10-180-552-98	Sequence 98, Appl
6	3.1	472	9	US-10-180-557-98	Sequence 98, Appl
6	3.1	472	9	US-10-173-700-98	Sequence 98, Appl
6	3.1	472	9	US-10-174-572-98	Sequence 98, Appl
6	3.1	472	9	US-10-174-579-98	Sequence 98, Appl
6	3.1	472	9	US-10-174-582-98	Sequence 98, Appl
6	3.1	472	9	US-10-174-588-98	Sequence 98, Appl
6	3.1	472	9	US-10-175-739-98	Sequence 98, Appl
6	3.1	472	9	US-10-175-740-98	Sequence 98, Appl
6	3.1	472	9	US-10-175-743-98	Sequence 98, Appl
6	3.1	472	9	US-10-176-488-98	Sequence 98, Appl
6	3.1	472	9	US-10-176-492-98	Sequence 98, Appl
6	3.1	472	9	US-10-176-747-98	Sequence 98, Appl
6	3.1	472	9	US-10-176-750-98	Sequence 98, Appl
6	3.1	472	9	US-10-176-985-98	Sequence 98, Appl
6	3.1	472	9	US-10-176-987-98	Sequence 98, Appl
6	3.1	472	9	US-10-176-991-98	Sequence 98, Appl
6	3.1	472	9	US-10-176-992-98	Sequence 98, Appl
6	3.1	472	9	US-10-176-993-98	Sequence 98, Appl

TELEPHONE: 206-622-4900
TELEFAX: 206-682-6031
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 196 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-106-582-58

Query Match 3.1%; Score 6; DB 4; Length 196;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 VKSDDI 32
|||||
Db 138 VKSDDI 143

RESULT 38

US-09-134-001C-3459
Sequence 3459, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
PRIOR FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 3459
LENGTH: 198
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3459

Query Match 3.1%; Score 6; DB 4; Length 198;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 FENIVK 28
|||||
Db 25 FENIVK 30

RESULT 39

US-09-134-001C-4784
Sequence 4784, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
PRIOR FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 4784
LENGTH: 244
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4784

Query Match 3.1%; Score 6; DB 4; Length 244;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 147 SOSNAN 152
|||||
Db 229 SOSNAN 234

RESULT 40

US-08-252-995D-11
Sequence 11, Application US/08252995D
Patent No. 5650501
GENERAL INFORMATION:
APPLICANT: Dennis, James W
APPLICANT: Heffernan, Mike
APPLICANT: Fode, Carol
TITLE OF INVENTION: NOVEL SERINE/THREONINE KINASE
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: BERESKIN & PARR
STREET: 40 King Street West
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5H 3Y2
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/252,995D
FILING DATE: 02-JUN-1994
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Kurdydyk, Linda M
REGISTRATION NUMBER: 34,971
REFERENCE/DOCKET NUMBER: 3153-96
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 364-7311
TELEFAX: (416) 361-1398
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 271 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Drosophila melanogaster
US-08-252-995D-11

Query Match 3.1%; Score 6; DB 1; Length 271;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 HKRRKS 131
|||||
Db 96 HKRRKS 101

Search completed: February 26, 2003, 10:02:05
Job time : 31 secs

RESULT 34
US-09-295-028-32
Sequence 32, Application US/09295028
Patent No. 6277381
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
APPLICANT: McNeill, Patricia D.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
AND TREATMENT OF EHRlichia INFECTION
FILE REFERENCE: 210121.439C4
CURRENT APPLICATION NUMBER: US/09/295,028
CURRENT FILING DATE: 1999-04-20
NUMBER OF SEQ ID NOS: 85
SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 32
LENGTH: 196
TYPE: PRT
ORGANISM: Ehrlichia sp.
US-09-295-028-32

Query Match 3.1%, Score 6; DB 4; Length 196;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 VKSDDI 32
|||||

Db 138 VKSDDI 143

RESULT 35
US-09-295-028-58
Sequence 58, Application US/09295028
Patent No. 6277381
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
APPLICANT: McNeill, Patricia D.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
AND TREATMENT OF EHRlichia INFECTION
FILE REFERENCE: 210121.439C4
CURRENT APPLICATION NUMBER: US/09/295,028
CURRENT FILING DATE: 1999-04-20
NUMBER OF SEQ ID NOS: 85
SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 58
LENGTH: 196
TYPE: PRT
ORGANISM: Ehrlichia sp.
US-09-295-028-58

Query Match 3.1%, Score 6; DB 4; Length 196;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 VKSDDI 32
|||||

Db 138 VKSDDI 143

RESULT 36
US-09-106-582-32
Sequence 32, Application US/09106582
Patent No. 6306402
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
THERAPY
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/106,582
FILING DATE: 29-JUN-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.439C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-622-4900
TELEFAX: 206-682-6031
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 196 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Ehrlichia
US-09-106-582-32

Query Match 3.1%, Score 6; DB 4; Length 196;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 VKSDDI 32
|||||

Db 138 VKSDDI 143

RESULT 37
US-09-106-582-58
Sequence 58, Application US/09106582
Patent No. 6306402
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
THERAPY
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/106,582
FILING DATE: 29-JUN-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.439C2
TELECOMMUNICATION INFORMATION:

ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/975,762
APPLICATION NUMBER: US/08/975,762
FILING DATE: 21-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.439
TELEPHONE: 206-622-4900
TELEFAX: 206-622-4900
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 196 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE: Ehrlichia
ORGANISM: Ehrlichia
US-08-975-762-32

Query Match 3.1%; Score 6; DB 4; Length 196;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 VKSDDI 32
Db 138 VKSDDI 143

RESULT 32
US-08-975-762-58
Sequence 58, Application US/08975762
Patent No. 6207169
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/975,762
FILING DATE: 21-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.439
TELECOMMUNICATION INFORMATION:

TREATMEN

TELEPHONE: 206-622-4900
TELEFAX: 206-682-6031
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 196 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-975-762-58

Query Match 3.1%; Score 6; DB 4; Length 196;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 VKSDDI 32
Db 138 VKSDDI 143

RESULT 33
US-08-821-324-32
Sequence 32, Application US/08821324
Patent No. 6231869
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/821,324
APPLICATION NUMBER: US/08/821,324
FILING DATE: 21-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.439
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-622-4900
TELEFAX: 206-682-6031
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 196 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE: Ehrlichia
ORGANISM: Ehrlichia
US-08-821-324-32

Query Match 3.1%; Score 6; DB 4; Length 196;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 VKSDDI 32
Db 138 VKSDDI 143

ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/602.010A
FILING DATE: February 15, 1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-46
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/863-9700
TELEFAX: 303/863-0223
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 149 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-602-010A-18

Query Match 3.1%; Score 6; DB 1; Length 149;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 156 TSKRRK 161
|||||
Db 109 TSKRRK 114

RESULT 29

US-08-680-726A-18
Sequence 18, Application US/08680726A
Patent No. 5804197
GENERAL INFORMATION:
APPLICANT: Haanes, Elizabeth J.
APPLICANT: Frank, Rexann S.
TITLE OF INVENTION: RECOMBINANT CANINE HERPESVIRUSES
NUMBER OF SEQUENCES: 92
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/680.726A
FILING DATE: 12-JUL-1996
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-46-C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 149 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-680-726A-18

Query Match 3.1%; Score 6; DB 1; Length 149;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 156 TSKRRK 161
|||||
Db 109 TSKRRK 114

RESULT 30

US-09-092-409-18
Sequence 18, Application US/09092409
Patent No. 6151478
GENERAL INFORMATION:
APPLICANT: Haanes, Elizabeth J.
APPLICANT: Frank, Rexann S.
TITLE OF INVENTION: RECOMBINANT CANINE HERPESVIRUSES
NUMBER OF SEQUENCES: 92
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/092.409
FILING DATE:
CLASSIFICATION:

PRIOR APPLICATION DATA:

PRIOR APPLICATION NUMBER: 08/680,726
FILING DATE: 12-JUL-1996
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-46-C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 149 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-092-409-18

Query Match 3.1%; Score 6; DB 4; Length 149;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 156 TSKRRK 161
|||||
Db 109 TSKRRK 114

RESULT 31

US-08-975-762-35
Sequence 32, Application US/08975762
Patent No. 6207169
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
CORRESPONDENCE ADDRESS:
NUMBER OF SEQUENCES: 73

US-08-680-726A-18

APPLICATION NUMBER: US/09/092,409
FILING DATE: 12-JUL-1996
CLASSIFICATION: 3.1; Score 6; DB 4; Length 119;
Prior Application Number: 08/680,726
Filing Date: 12-JUL-1996
Attorney/Agent Information:
Name: Connell, Gary J.
Registration Number: 32,020
Reference/Docket Number: 2618-46-C1
Telecommunication Information:
Telephone: (303) 863-9700
Telefax: (303) 863-0223
Information for Seq ID No: 62;
Sequence Characteristics:
Length: 119 amino acids
Type: amino acid
Topology: linear
Molecule Type: protein
US-09-092-409-62

Query Match 3.1; Score 6; DB 4; Length 119;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 156 TSKRRK 161
Db 79 TSKRRK 84

RESULT 26
US-08-680-726A-66
Sequence 66, Application US/08680726A
Patent No. 5804197
GENERAL INFORMATION:
APPLICANT: Haanes, Elizabeth J.
APPLICANT: Frank, Rexann S.
TITLE OF INVENTION: RECOMBINANT CANINE HERPESVIRUSES
NUMBER OF SEQUENCES: 92
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/680,726A
FILING DATE: 12-JUL-1996
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-46-C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 66;
SEQUENCE CHARACTERISTICS:
LENGTH: 139 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-680-726A-66

Query Match 3.1; Score 6; DB 1; Length 139;
Best Local Similarity 100.0%; Pred. No. 96;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 RSKSQS 137
Db 120 RSKSQS 125

RESULT 27
US-09-092-409-66
Sequence 66, Application US/09092409
Patent No. 6159478
GENERAL INFORMATION:
APPLICANT: Haanes, Elizabeth J.
APPLICANT: Frank, Rexann S.
TITLE OF INVENTION: RECOMBINANT CANINE HERPESVIRUSES
NUMBER OF SEQUENCES: 92
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/092,409
FILING DATE:
CLASSIFICATION:
Prior Application Data:
APPLICATION NUMBER: 08/680,726
FILING DATE: 12-JUL-1996
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-46-C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 66;
SEQUENCE CHARACTERISTICS:
LENGTH: 139 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-092-409-66

Query Match 3.1; Score 6; DB 4; Length 139;
Best Local Similarity 100.0%; Pred. No. 96;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 RSKSQS 137
Db 120 RSKSQS 125

RESULT 28
US-08-602-010A-18
Sequence 18, Application US/08602010A
Patent No. 5753235
GENERAL INFORMATION:
APPLICANT: Haanes, Elizabeth J.
APPLICANT: Frank, Rexann S.
TITLE OF INVENTION: RECOMBINANT CANINE HERPESVIRUSES
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.

REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454312-2074
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 108 amino acids
TYPE: amino acid
STRANDEDNESS: n/a
TOPOLOGY: linear
MOLECULE TYPE: amino acid
US-08-710-749-25

Query Match 3.1%; Score 6; DB 2; Length 108;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 94 EDYKE 99
|||||
Db 68 EDYKE 73

RESULT 23
US-08-710-749-26
Sequence 26, Application US/08710749
Patent No. 5955089
GENERAL INFORMATION:
APPLICANT: Briles, David E.
APPLICANT: Hollingshead, Susan
APPLICANT: Becker, Robert
TITLE OF INVENTION: STRAIN SELECTION OF PNEUMOCOCCAL SURFACE
TITLE OF INVENTION: PROTEINS
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: Curtis, Morris & Safford
STREET: 530 Fifth Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/710,749
FILING DATE: 20-SEP-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Frommer, William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454312-2074
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 108 amino acids
TYPE: amino acid
STRANDEDNESS: n/a
TOPOLOGY: linear
MOLECULE TYPE: amino acid
US-08-710-749-26

Query Match 3.1%; Score 6; DB 2; Length 108;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 94 EDYKE 99
|||||
Db 68 EDYKE 73

RESULT 24
US-08-680-726A-6;
Sequence 62, Application US/08680726A
Patent No. 5804197
GENERAL INFORMATION:
APPLICANT: Haanes, Elizabeth J.
APPLICANT: Frank, Rexann S.
TITLE OF INVENTION: RECOMBINANT CANINE HERPESVIRUSES
NUMBER OF SEQUENCES: 92
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/680,726A
FILING DATE: 12-JUL-1996
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-46-C1
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 62:
SEQUENCE CHARACTERISTICS:
LENGTH: 119 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-680-726A-62

Query Match 3.1%; Score 6; DB 1; Length 119;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 156 TSKRRK 161
|||||
Db 79 TSKRRK 84

RESULT 25
US-09-092-409-6;
Sequence 62, Application US/09092409
Patent No. 6159478
GENERAL INFORMATION:
APPLICANT: Haanes, Elizabeth J.
APPLICANT: Frank, Rexann S.
TITLE OF INVENTION: RECOMBINANT CANINE HERPESVIRUSES
NUMBER OF SEQUENCES: 92
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

```
;
; TOPOLOGY: linear
; MOLECULE TYPE: amino acid
US-08-710-749-22

Query Match          3.1%; Score 6; DB 2; Length 108;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 EDYIKE 99
Db 68 EDYIKE 73

RESULT 20
US-08-710-749-23
; Sequence 23, Application US/08710749
; Patent No. 5955089
; GENERAL INFORMATION:
; APPLICANT: Briles, David E.
; APPLICANT: Hollingshead, Susan
; APPLICANT: Becker, Robert
; TITLE OF INVENTION: STRAIN SELECTION OF PNEUMOCOCCAL SURFACE
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/710,749
; FILING DATE: 20-SEP-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454312-2074
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 108 amino acids
; TYPE: amino acid
; STRANDEDNESS: n/a
; TOPOLOGY: linear
; MOLECULE TYPE: amino acid
US-08-710-749-24

Query Match          3.1%; Score 6; DB 2; Length 108;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 EDYIKE 99
Db 68 EDYIKE 73

RESULT 22
US-08-710-749-25
; Sequence 25, Application US/08710749
; Patent No. 5955089
; GENERAL INFORMATION:
; APPLICANT: Briles, David E.
; APPLICANT: Hollingshead, Susan
; APPLICANT: Becker, Robert
; TITLE OF INVENTION: STRAIN SELECTION OF PNEUMOCOCCAL SURFACE
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/710,749
; FILING DATE: 20-SEP-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454312-2074
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 108 amino acids
; TYPE: amino acid
; STRANDEDNESS: n/a
; TOPOLOGY: linear
; MOLECULE TYPE: amino acid
US-08-710-749-23

Query Match          3.1%; Score 6; DB 2; Length 108;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 EDYIKE 99
Db 68 EDYIKE 73

RESULT 21
US-08-710-749-24
; Sequence 24, Application US/08710749
; Patent No. 5955089
; GENERAL INFORMATION:
; APPLICANT: Briles, David E.
; APPLICANT: Hollingshead, Susan
; APPLICANT: Becker, Robert
```

TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 213:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
PCT-US93-08528-213

Query Match 3.1% Score 6; DB 5; Length 38;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 FIVGDW 67
Db 19 FIVGDW 24

RESULT 17

US-08-710-749-6
Sequence 6, Application US/08710749

Patent No. 5955089

GENERAL INFORMATION:

APPLICANT: Briles, David E.

APPLICANT: Hollingshead, Susan

APPLICANT: Becker, Robert

TITLE OF INVENTION: STRAIN SELECTION OF PNEUMOCOCCAL SURFACE

TITLE OF INVENTION: PROTEINS

NUMBER OF SEQUENCES: 28

CORRESPONDENCE ADDRESS:

ADDRESSEE: Curtis, Morris & Safford

STREET: 530 Fifth Avenue

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/710,749

FILING DATE: 20-SEP-1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Frommer, William S.

REGISTRATION NUMBER: 25,506

REFERENCE/DOCKET NUMBER: 454312-2074

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 840-3333

TELEFAX: (212) 840-0712

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 101 amino acids

TYPE: amino acid

STRANDEDNESS: n/a

TOPOLOGY: linear

MOLECULE TYPE: amino acid

US-08-710-749-6

Query Match 3.1% Score 6; DB 2; Length 101;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 EDYIKE 99
Db 10 EDYIKE 15

RESULT 18

US-08-407-165-3

Sequence 3, Application US/08407165

Patent No. 6054280

GENERAL INFORMATION:

APPLICANT: LENNON, MARK A.

APPLICANT: FEIGUSON, KATHRYN M.

APPLICANT: SICHER, PAUL B.

APPLICANT: SCHLESSINGER, JOSEPH

TITLE OF INVENTION: METHODS FOR DIAGNOSIS AND TREATMENT OF

TITLE OF INVENTION: PH DOMAIN SIGNAL TRANSDUCTION DISORDERS

FILE REFERENCE: 211/156

CURRENT APPLICATION NUMBER: US/08/407,165

CURRENT FILING DATE: 1995-03-20

NUMBER OF SEQ ID NOS: 4

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 3

LENGTH: 105

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

OTHER INFORMATION: Spectrin

US-08-407-165-3

Query Match 3.1% Score 6; DB 3; Length 105;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 GFYKDA 81
Db 36 GFYKDA 41

RESULT 19

US-08-710-749-22

Sequence 22, Application US/08710749

Patent No. 5955089

GENERAL INFORMATION:

APPLICANT: Briles, David E.

APPLICANT: Hollingshead, Susan

APPLICANT: Becker, Robert

TITLE OF INVENTION: STRAIN SELECTION OF PNEUMOCOCCAL SURFACE

TITLE OF INVENTION: PROTEINS

NUMBER OF SEQUENCES: 28

CORRESPONDENCE ADDRESS:

ADDRESSEE: Curtis, Morris & Safford

STREET: 530 Fifth Avenue

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/710,749

FILING DATE: 20-SEP-1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Frommer, William S.

REGISTRATION NUMBER: 25,506

REFERENCE/DOCKET NUMBER: 454312-2074

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 840-3333

TELEFAX: (212) 840-0712

INFORMATION FOR SEQ ID NO: 22:

SEQUENCE CHARACTERISTICS:

LENGTH: 108 amino acids

TYPE: amino acid

STRANDEDNESS: n/a

```
RESULT 13
5240706-21
; PATENT NO. 5240706
; APPLICANT: FAULDS, DARYL
; TITLE OF INVENTION: INTRANASSAL ADMINISTRATION OF MYCOPLASMA
; HYOPNEUMONIAE ANTIGEN
; NUMBER OF SEQUENCES: 22
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/334,586
; FILING DATE: 07-APR-1989
; SEQ ID NO: 21
; LENGTH: 24
5240706-21

Query Match          3.1%  Score 6;  DB 6;  Length 24;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 109 VLENPN 114
      |||||
Db 18 VLENPN 23

RESULT 14
US-09-177-249-217
; Sequence 217, Application US/09177249
; Patent No. 6229064
; GENERAL INFORMATION:
; APPLICANT: FISCHER, ROBERT L.
; APPLICANT: OHAD, NIR
; APPLICANT: KIYOSUE, TOMOHIRO
; APPLICANT: YADAGARI, RAMIN
; APPLICANT: MARGOSSIAN, LINDA
; APPLICANT: HARADA, JOHN
; APPLICANT: GOLDBERG, ROBERT B.
; TITLE OF INVENTION: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
; TITLE OF INVENTION: NUCLEIC ACIDS THAT CONTROL SEED AND FRUIT
; FILE REFERENCE: 023070-086120US
; CURRENT APPLICATION NUMBER: US/09/177,249
; CURRENT FILING DATE: 1998-10-22
; EARLIER APPLICATION NUMBER: US 09/071,838
; EARLIER FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 324
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 217
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Arabidopsis sp.
US-09-177-249-217

Query Match          3.1%  Score 6;  DB 4;  Length 25;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 KNSRIS 91
      |||||
Db 9 KNSRIS 14

RESULT 15
US-08-118-270-213
; Sequence 213, Application US/08118270
; Patent No. 5508384
; GENERAL INFORMATION:
; APPLICANT: MURPHY, RANDALL B.
; APPLICANT: SCHUSTER, DAVID I.
; TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
; TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
; NUMBER OF SEQUENCES: 348
; CORRESPONDENCE ADDRESS:
```

```
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/118,270
FILING DATE: 09-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/943,236
FILING DATE: 10-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Townsend, Kevin G.
REGISTRATION NUMBER: 34,033
REFERENCE/DOCKET NUMBER: MURPHY-2A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 213:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-118-270-213

Query Match          3.1%  Score 6;  DB 1;  Length 38;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 FIVGDW 67
      |||||
Db 19 FIVGDW 24

RESULT 16
PCT-US93-08528-213
; Sequence 213, Application PC/TUS9308528
; GENERAL INFORMATION:
; APPLICANT: New York University
; TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
; TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
; NUMBER OF SEQUENCES: 348
; CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/08528
FILING DATE: 09-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/943,236
FILING DATE: 10-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Townsend, Kevin G.
REGISTRATION NUMBER: 34,033
REFERENCE/DOCKET NUMBER: MURPHY-2 PCT
```

Patent No. 5858701
GENERAL INFORMATION:
APPLICANT: WHITE, Morris F.
APPLICANT: SUN, Xiao Jian
APPLICANT: PIERCE, Jacalyn H.
TITLE OF INVENTION: THE IRS FAMILY OF GENES
NUMBER OF SEQUENCES: 64
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/317,310A
FILING DATE: 03-OCT-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Louis Myers
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: JDP-022
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 1321 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-317-310A-16

Query Match 3.6%; Score 7; DB 2;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 132 RSKSQSS 138
|||||||
Db 300 RSKSQSS 306

RESULT 11

US-08-317-310A-64
Sequence 64, Application US/08317310A
Patent No. 5858701
GENERAL INFORMATION:
APPLICANT: WHITE, Morris F.
APPLICANT: SUN, Xiao Jian
APPLICANT: PIERCE, Jacalyn H.
TITLE OF INVENTION: THE IRS FAMILY OF GENES
NUMBER OF SEQUENCES: 64
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/317,310A
FILING DATE: 03-OCT-1994
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Louis Myers
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: JDP-022
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 64:
SEQUENCE CHARACTERISTICS:
LENGTH: 1321 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-317-310A-64

Query Match 3.6%; Score 7; DB 2;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 132 RSKSQSS 138
|||||||
Db 300 RSKSQSS 306

RESULT 12

PCT-US95-13041-16
Sequence 16, Application PC/TUS9513041
GENERAL INFORMATION:
APPLICANT: WHITE, Morris F.
APPLICANT: SUN, Xiao Jian
APPLICANT: PIERCE, Jacalyn H.
TITLE OF INVENTION: THE IRS FAMILY OF GENES
NUMBER OF SEQUENCES: 63
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 30 State Street, Suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/13041
FILING DATE: Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/317,310
FILING DATE: 03-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: Louis Myers
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: JDP-022PC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 1321 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-13041-16

Query Match 3.6%; Score 7; DB 5;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 132 RSKSQSS 138
|||||||
Db 300 RSKSQSS 306

US-09-085-761A-2
; Sequence 2, Application US/09085761A
; Patent No. 6335178
; GENERAL INFORMATION:
; APPLICANT: Weiner, Joel H.
; APPLICANT: Turner, Raymond J.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PROTEIN
; TITLE OF INVENTION: SECRETION
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/085,761A
; FILING DATE: 28-MAY-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Carroll, Peter G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: UALB-03356
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 284 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-09-085-761A-2

Query Match 3.6%; Score 7; DB 4; Length 284;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 133 SKSOSSK 139
Db 275 SKSOSSK 281

RESULT 5
US-08-937-067-6
; Sequence 6, Application US/08937067
; Patent No. 6433155
; GENERAL INFORMATION:
; APPLICANT: Umansky, Samuil
; APPLICANT: Melkonyan, Hovsep
; TITLE OF INVENTION: A FAMILY OF GENES ENCODING
; TITLE OF INVENTION: APOPTOSIS-RELATED PEPTIDES; PEPTIDES ENCODED THEREBY AND
; TITLE OF INVENTION: METHODS OF USE THEREOF
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,067
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Lehnhardt, Susan K.
; REGISTRATION NUMBER: 33,943
; REFERENCE/DOCKET NUMBER: 23647-20018.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 813-5600
; TELEFAX: (650) 494-0792
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 317 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-937-067-6

Query Match 3.6%; Score 7; DB 4; Length 317;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 159 KRKDTKF 165
Db 236 KRKDTKF 242

RESULT 6
US-08-317-310A-15
; Sequence 15, Application US/08317310A
; Patent No. 5858701
; GENERAL INFORMATION:
; APPLICANT: WHITE, MORRIS F.
; APPLICANT: SUN, Xiao Jian
; APPLICANT: PIERCE, Jacalyn H.
; TITLE OF INVENTION: THE IRS FAMILY OF GENES
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 78 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/317,310A
; FILING DATE: 03-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Louis Myers
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: JDP-022
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1234 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
US-08-317-310A-15

Query Match 3.6%; Score 7; DB 2; Length 1234;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
977 5 2.6 335 4 US-09-232-191-21 Sequence 21, Appl
978 5 2.6 335 4 US-09-232-200-21 Sequence 21, Appl
979 5 2.6 335 4 US-09-232-197-21 Sequence 21, Appl
980 5 2.6 335 4 US-09-180-100-20 Sequence 20, Appl
981 5 2.6 335 4 US-09-232-201-21 Sequence 21, Appl
982 5 2.6 335 4 US-09-565-918-3 Sequence 3, Appl
983 5 2.6 335 5 PCT-US95-17083-2 Sequence 7, Appl
984 5 2.6 336 1 US-07-928-462-2 Sequence 2, Appl
985 5 2.6 336 4 US-08-273-247-2 Sequence 2, Appl
986 5 2.6 337 1 US-07-966-187-25 Sequence 25, Appl
987 5 2.6 337 2 US-08-466-033-162 Sequence 162, App
988 5 2.6 337 2 US-08-464-134-162 Sequence 162, App
989 5 2.6 337 2 US-08-461-361-162 Sequence 162, App
990 5 2.6 337 3 US-09-032-372-2 Sequence 2, Appl
991 5 2.6 337 3 US-08-686-528A-2 Sequence 2, Appl
992 5 2.6 337 3 US-09-190-965-1 Sequence 1, Appl
993 5 2.6 337 4 US-09-456-287-2 Sequence 2, Appl
994 5 2.6 337 4 US-09-470-253-1 Sequence 1, Appl
995 5 2.6 338 1 US-07-966-187-3 Sequence 3, Appl
996 5 2.6 338 1 US-07-966-187-16 Sequence 16, Appl
997 5 2.6 338 4 US-08-448-398-13 Sequence 13, Appl
1000

ALIGNMENTS

US-09-134-001C-4894
; Sequence 4894, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4894
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4894

Query Match 3.6%; Score 7; DB 4; Length 132;
Best Local Similarity 100.0%; Pred. No. 8.2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 EYCNFGC 105
Db 78 EYCNFGC 84

RESULT 2
US-09-297-937C-7
; Sequence 7, Application US/09297937C
; Patent No. 6337199
; GENERAL INFORMATION:
; APPLICANT: YUN, Do Young
; APPLICANT: PAN, Jae Gu
; TITLE OF INVENTION: Membrane-Bound Gluconate Dehydrogenase, Gene Sequence
; TITLE OF INVENTION: Encoding the Same and Production of 2-Keto-D-Gluconate
; FILE REFERENCE: P66159US0
; CURRENT APPLICATION NUMBER: US/09/297,937C
; CURRENT FILING DATE: 1999-05-11
```

```
; PRIOR APPLICATION NUMBER: PCT/KR98/00296
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: KR 97-48802
; PRIOR FILING DATE: 1997-09-25
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 7
; LENGTH: 220
; TYPE: PRT
; ORGANISM: Erwinia cyripedii
US-09-297-937C-7

Query Match 3.6%; Score 7; DB 4; Length 220;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 136 QSSKSQT 142
Db 45 QSSKSQT 51

RESULT 3
US-09-053-197A-2
; Sequence 2, Application US/09053197A
; Patent No. 6022952
; GENERAL INFORMATION:
; APPLICANT: Weiner, Joel H.
; APPLICANT: Turner, Raymond J.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PROTEIN
; TITLE OF INVENTION: SECRETION
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/09/053,197A
; FILING DATE: 01-APR-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Macknight, Kamrin T.
; REGISTRATION NUMBER: 38,230
; REFERENCE/DOCKET NUMBER: UALB-03293
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 284 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-09-053-197A-2

Query Match 3.6%; Score 7; DB 3; Length 284;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 SKSQSSK 139
Db 275 SKSQSSK 281

RESULT 4
```

831	5	2.6	281	4	US-09-479-524-3	Sequence 3, Appli	904	5	2.6	310	3	US-09-063-869-4	Sequence 4, Appli
832	5	2.6	281	4	US-08-339-214-8	Sequence 8, Appli	905	5	2.6	313	3	US-08-686-528A-3	Sequence 3, Appli
833	5	2.6	281	4	US-08-339-214-30	Sequence 30, Appli	906	5	2.6	313	4	US-09-456-287-3	Sequence 3, Appli
834	5	2.6	281	4	US-09-589-287B-6	Sequence 6, Appli	907	5	2.6	314	1	US-08-444-231-19	Sequence 19, Appli
835	5	2.6	281	4	US-09-157-864-10	Sequence 10, Appli	908	5	2.6	314	1	US-08-152-443A-19	Sequence 19, Appli
836	5	2.6	281	4	US-09-006-755B-1	Sequence 1, Appli	909	5	2.6	314	5	PCT-US95-17083-4	Sequence 4, Appli
837	5	2.6	281	5	PCT-US95-00362-2	Sequence 2, Appli	910	5	2.6	315	3	US-09-135-639-4	Sequence 4, Appli
838	5	2.6	282	1	US-07-712-476A-5	Sequence 5, Appli	911	5	2.6	315	4	US-09-134-001C-4380	Sequence 4380, Ap
839	5	2.6	283	1	US-08-583-672-2	Sequence 2, Appli	912	5	2.6	316	2	US-08-846-762-4	Sequence 4, Appli
840	5	2.6	283	2	US-08-202-044-2	Sequence 2, Appli	913	5	2.6	317	2	US-08-973-275-1	Sequence 1, Appli
841	5	2.6	283	4	US-08-751-344B-2	Sequence 2, Appli	914	5	2.6	318	3	US-08-859-167-4	Sequence 4, Appli
842	5	2.6	284	2	US-08-320-148B-2	Sequence 2, Appli	915	5	2.6	318	3	US-08-859-167-6	Sequence 6, Appli
843	5	2.6	284	3	US-08-589-028-6	Sequence 6, Appli	916	5	2.6	318	3	US-09-109-273-4	Sequence 4, Appli
844	5	2.6	284	3	US-08-491-954-3	Sequence 3, Appli	917	5	2.6	318	3	US-09-109-273-6	Sequence 6, Appli
845	5	2.6	284	3	US-08-784-582-6	Sequence 6, Appli	918	5	2.6	318	4	US-08-949-246-2	Sequence 2, Appli
846	5	2.6	284	4	US-08-785-271-6	Sequence 6, Appli	919	5	2.6	318	4	US-09-276-993-4	Sequence 4, Appli
847	5	2.6	284	4	US-09-031-898-2	Sequence 2, Appli	920	5	2.6	318	4	US-09-276-993-6	Sequence 6, Appli
848	5	2.6	285	2	US-08-418-071-17	Sequence 17, Appli	921	5	2.6	318	4	US-09-134-001C-3603	Sequence 3603, Ap
849	5	2.6	286	5	PCT-US92-00282-9	Sequence 9, Appli	922	5	2.6	320	1	US-08-565-386-17	Sequence 17, Appli
850	5	2.6	287	3	US-08-815-190A-16	Sequence 16, Appli	923	5	2.6	321	4	US-08-506-296B-54	Sequence 54, Appli
851	5	2.6	287	3	US-08-549-515-10	Sequence 10, Appli	924	5	2.6	321	4	US-09-134-001C-2890	Sequence 2890, Ap
852	5	2.6	287	4	US-08-862-124-17	Sequence 17, Appli	925	5	2.6	322	1	US-08-097-828-1	Sequence 1, Appli
853	5	2.6	289	5	PCT-US92-00282-13	Sequence 13, Appli	926	5	2.6	322	1	US-08-036-210-11	Sequence 11, Appli
854	5	2.6	289	5	PCT-US92-00282-15	Sequence 15, Appli	927	5	2.6	322	1	US-08-480-756-1	Sequence 1, Appli
855	5	2.6	290	2	US-08-903-801-1	Sequence 1, Appli	928	5	2.6	322	2	US-08-462-403-1	Sequence 1, Appli
856	5	2.6	290	4	US-09-295-055-1	Sequence 4, Appli	929	5	2.6	322	2	US-08-286-819A-2	Sequence 2, Appli
857	5	2.6	292	2	US-08-701-191A-40	Sequence 40, Appli	930	5	2.6	322	2	US-08-449-609-11	Sequence 11, Appli
858	5	2.6	294	2	US-08-424-641B-6	Sequence 6, Appli	931	5	2.6	322	3	US-08-980-357-2	Sequence 2, Appli
859	5	2.6	294	2	US-08-820-980-6	Sequence 6, Appli	932	5	2.6	322	3	PCT-US92-00331-1	Sequence 1, Appli
860	5	2.6	294	2	US-08-826-439-6	Sequence 6, Appli	933	5	2.6	322	5	PCT-US93-10419-1	Sequence 1, Appli
861	5	2.6	294	4	US-09-131-028A-9	Sequence 9, Appli	934	5	2.6	326	2	US-08-986-217-3	Sequence 29, Appli
862	5	2.6	294	4	US-09-131-028A-11	Sequence 11, Appli	935	5	2.6	326	4	US-09-066-046-29	Sequence 29, Appli
863	5	2.6	295	1	US-08-149-091-1	Sequence 1, Appli	936	5	2.6	326	4	US-09-066-047-15	Sequence 15, Appli
864	5	2.6	295	2	US-08-637-761-1	Sequence 1, Appli	937	5	2.6	327	1	US-08-080-386-2	Sequence 2, Appli
865	5	2.6	295	4	US-09-134-001C-5643	Sequence 5643, Ap	938	5	2.6	327	2	US-08-390-000A-2	Sequence 2, Appli
866	5	2.6	295	5	PCT-US94-12672-1	Sequence 1, Appli	939	5	2.6	327	4	US-09-446-504-3	Sequence 3, Appli
867	5	2.6	296	1	US-07-712-476A-1	Sequence 1, Appli	940	5	2.6	327	4	US-09-712-266-3	Sequence 3, Appli
868	5	2.6	296	4	US-08-944-604-20	Sequence 20, Appli	941	5	2.6	327	4	US-09-583-492-10	Sequence 10, Appli
869	5	2.6	297	2	US-08-874-347-23	Sequence 23, Appli	942	5	2.6	328	1	US-08-423-691-2	Sequence 2, Appli
870	5	2.6	297	2	US-09-006-535-4	Sequence 4, Appli	943	5	2.6	328	1	US-08-080-386-4	Sequence 4, Appli
871	5	2.6	297	3	US-09-093-522-23	Sequence 23, Appli	944	5	2.6	328	2	US-08-390-000A-4	Sequence 4, Appli
872	5	2.6	298	1	US-08-149-091-6	Sequence 6, Appli	945	5	2.6	328	3	US-08-867-260-2	Sequence 2, Appli
873	5	2.6	298	2	US-08-637-761-6	Sequence 6, Appli	946	5	2.6	328	4	US-09-501-192-9	Sequence 9, Appli
874	5	2.6	298	2	US-09-006-535-3	Sequence 3, Appli	947	5	2.6	328	4	US-09-428-747-4	Sequence 4, Appli
875	5	2.6	298	5	PCT-US94-12672-6	Sequence 6, Appli	948	5	2.6	328	4	US-09-615-192A-274	Sequence 274, App
876	5	2.6	299	4	US-09-134-001C-5227	Sequence 5227, Ap	949	5	2.6	329	1	US-07-991-587A-7	Sequence 7, Appli
877	5	2.6	300	2	US-08-828-010-2	Sequence 2, Appli	950	5	2.6	329	1	US-08-309-985-7	Sequence 7, Appli
878	5	2.6	300	6	5340934-6	Patent No. 5340934	951	5	2.6	329	4	US-09-189-527-4	Sequence 4, Appli
879	5	2.6	301	4	US-08-944-604-18	Sequence 18, Appli	952	5	2.6	330	2	US-08-525-864A-4	Sequence 4, Appli
880	5	2.6	301	4	US-09-288-143-98	Sequence 98, Appli	953	5	2.6	331	1	US-08-466-033-166	Sequence 166, App
881	5	2.6	302	4	US-09-457-046B-18	Sequence 18, Appli	954	5	2.6	331	2	US-08-444-733-166	Sequence 166, App
882	5	2.6	303	1	US-08-109-391A-2	Sequence 2, Appli	955	5	2.6	331	2	US-08-464-134-166	Sequence 166, App
883	5	2.6	303	1	US-08-459-019A-2	Sequence 2, Appli	956	5	2.6	331	2	US-08-461-361-166	Sequence 166, App
884	5	2.6	303	2	US-08-460-428A-2	Sequence 2, Appli	957	5	2.6	331	2	US-08-485-910-166	Sequence 166, App
885	5	2.6	303	3	US-08-458-860A-2	Sequence 2, Appli	958	5	2.6	331	2	US-08-986-217-7	Sequence 7, Appli
886	5	2.6	303	4	US-09-105-390-38	Sequence 38, Appli	959	5	2.6	331	4	US-09-086-483A-3	Sequence 3, Appli
887	5	2.6	304	1	US-07-851-976B-8	Sequence 8, Appli	960	5	2.6	331	4	US-09-199-637A-333	Sequence 333, App
888	5	2.6	304	1	US-08-291-609-8	Sequence 8, Appli	961	5	2.6	332	2	US-08-446-875-12	Sequence 12, Appli
889	5	2.6	304	1	US-08-401-136-8	Sequence 8, Appli	962	5	2.6	332	2	US-08-102-385G-12	Sequence 12, Appli
890	5	2.6	304	3	US-08-850-554-8	Sequence 8, Appli	963	5	2.6	332	4	US-09-105-390-54	Sequence 54, Appli
891	5	2.6	304	4	US-08-862-124-14	Sequence 14, Appli	964	5	2.6	333	4	US-08-961-083-54	Sequence 54, Appli
892	5	2.6	304	4	US-09-739-455-13	Sequence 13, Appli	965	5	2.6	334	1	US-08-118-270-73	Sequence 73, Appli
893	5	2.6	304	4	US-09-739-455-23	Sequence 23, Appli	966	5	2.6	334	1	US-08-036-210-7	Sequence 7, Ap211
894	5	2.6	305	4	US-09-134-001C-3192	Sequence 3192, Ap	967	5	2.6	334	2	US-08-449-609-7	Sequence 7, Ap211
895	5	2.6	305	4	US-09-451-900-1	Sequence 1, Appli	968	5	2.6	334	4	US-09-218-363-11	Sequence 11, Appli
896	5	2.6	306	4	US-09-921-318-1	Sequence 1, Appli	969	5	2.6	334	5	PCT-US93-08528-73	Sequence 73, Appli
897	5	2.6	307	4	US-08-949-246-4	Sequence 4, Appli	970	5	2.6	334	6	5290690-11	Patent No. 5290590
898	5	2.6	307	4	US-09-134-001C-5144	Sequence 5144, Ap	971	5	2.6	335	2	US-08-219-237B-2	Sequence 2, Appli
899	5	2.6	308	4	US-08-952-089A-13	Sequence 13, Appli	972	5	2.6	335	2	US-08-409-338-1	Sequence 1, Appli
900	5	2.6	309	1	US-07-966-187-1	Sequence 1, Appli	973	5	2.6	335	4	US-08-815-469-6	Sequence 6, Appli
901	5	2.6	309	3	US-08-605-284B-11	Sequence 11, Appli	974	5	2.6	335	4	US-09-230-640-2	Sequence 2, Appli
902	5	2.6	309	4	US-08-448-398-11	Sequence 11, Appli	975	5	2.6	335	4	US-09-006-353A-7	Sequence 7, Appli
903	5	2.6	310	2	US-08-695-355-4	Sequence 4, Appli	976	5	2.6	335	4	US-08-468-560C-2	Sequence 2, Appli

685	5	2.6	192	4	US-09-303-120B-8	Sequence 8, Appl	758	5	2.6	243	4	US-09-181-958-1	Sequence 1, Appl
686	5	2.6	192	4	US-09-820-576-8	Sequence 8, Appl	759	5	2.6	243	4	US-09-370-838-39	Sequence 39, Appl
687	5	2.6	192	4	US-09-218-363-18	Sequence 18, Appl	760	5	2.6	243	4	US-09-370-838-42	Sequence 42, Appl
688	5	2.6	198	3	US-08-737-248-6	Sequence 6, Appl	761	5	2.6	244	4	US-09-370-838-43	Sequence 43, Appl
689	5	2.6	199	2	US-08-405-175A-3	Sequence 3, Appl	762	5	2.6	244	4	US-09-370-838-46	Sequence 46, Appl
690	5	2.6	199	4	US-09-497-779A-6	Sequence 6, Appl	763	5	2.6	245	4	US-09-370-838-40	Sequence 40, Appl
691	5	2.6	200	2	US-08-405-175A-4	Sequence 4, Appl	764	5	2.6	246	4	US-08-822-774-45	Sequence 45, Appl
692	5	2.6	200	4	US-08-858-207A-390	Sequence 390, App	765	5	2.6	246	4	US-09-632-711-45	Sequence 45, Appl
693	5	2.6	201	2	US-09-933-750C-21	Sequence 21, Appl	766	5	2.6	246	4	US-09-632-703B-45	Sequence 45, Appl
694	5	2.6	201	4	US-09-234-613-21	Sequence 21, Appl	767	5	2.6	246	4	US-09-632-702-45	Sequence 45, Appl
695	5	2.6	202	4	US-09-370-838-38	Sequence 38, Appl	768	5	2.6	248	4	US-09-163-285-2	Sequence 2, Appl
696	5	2.6	203	4	US-09-134-001C-3111	Sequence 3111, Ap	769	5	2.6	248	4	US-09-134-001C-3731	Sequence 3731, Ap
697	5	2.6	204	2	US-08-516-801-3	Sequence 3, Appl	770	5	2.6	248	4	US-09-452-239-26	Sequence 26, Appl
698	5	2.6	204	4	US-08-248-355-3	Sequence 3, Appl	771	5	2.6	249	1	US-08-680-726A-88	Sequence 88, Appl
699	5	2.6	204	5	PCT-US95-06683-3	Sequence 3, Appl	772	5	2.6	249	4	US-09-092-409-88	Sequence 88, Appl
700	5	2.6	205	2	US-08-684-024-1	Sequence 1, Appl	773	5	2.6	249	4	US-09-578-303-2	Sequence 2, Appl
701	5	2.6	205	2	US-08-684-024-6	Sequence 6, Appl	774	5	2.6	251	2	US-08-401-068-12	Sequence 12, Appl
702	5	2.6	205	2	US-08-684-024-7	Sequence 7, Appl	775	5	2.6	251	2	US-08-846-338-12	Sequence 12, Appl
703	5	2.6	205	3	US-09-145-868-1	Sequence 1, Appl	776	5	2.6	251	3	US-08-411-768B-3	Sequence 3, Appl
704	5	2.6	205	3	US-09-145-868-6	Sequence 6, Appl	777	5	2.6	255	4	US-09-813-817-2	Sequence 2, Appl
705	5	2.6	205	3	US-09-145-868-7	Sequence 7, Appl	778	5	2.6	255	4	US-09-978-197-2	Sequence 2, Appl
706	5	2.6	208	4	US-09-134-001C-5600	Sequence 5600, Ap	779	5	2.6	258	1	US-08-149-091-2	Sequence 2, Appl
707	5	2.6	208	4	US-09-549-831-2	Sequence 2, Appl	780	5	2.6	258	2	US-08-637-761-2	Sequence 2, Appl
708	5	2.6	209	4	US-09-164-193-8	Sequence 8, Appl	781	5	2.6	258	4	US-09-497-779A-9	Sequence 9, Appl
709	5	2.6	209	4	US-09-221-448A-8	Sequence 8, Appl	782	5	2.6	258	4	US-09-134-001C-4806	Sequence 4806, Ap
710	5	2.6	209	6	5427917-2	Patent No. 5427917	783	5	2.6	258	4	US-09-134-001C-5593	Sequence 5593, Ap
711	5	2.6	210	2	US-08-781-560-4	Sequence 4, Appl	784	5	2.6	258	5	PCT-US94-12672-2	Sequence 2, Appl
712	5	2.6	210	2	US-08-933-750C-6	Sequence 6, Appl	785	5	2.6	259	2	US-08-402-804-8	Sequence 8, Appl
713	5	2.6	210	2	US-08-747-536-2	Sequence 2, Appl	786	5	2.6	260	2	US-08-763-919-2	Sequence 2, Appl
714	5	2.6	210	2	US-08-747-536-4	Sequence 4, Appl	787	5	2.6	260	4	US-08-560-729-2	Sequence 2, Appl
715	5	2.6	210	2	US-08-747-536-6	Sequence 6, Appl	788	5	2.6	260	4	US-09-721-259-2	Sequence 2, Appl
716	5	2.6	210	2	US-08-747-536-8	Sequence 8, Appl	789	5	2.6	261	2	US-08-989-478-16	Sequence 16, Appl
717	5	2.6	210	3	US-08-792-014-4	Sequence 4, Appl	790	5	2.6	261	3	US-08-996-685-16	Sequence 16, Appl
718	5	2.6	210	4	US-09-234-613-6	Sequence 6, Appl	791	5	2.6	261	4	US-09-097-889-19	Sequence 19, Appl
719	5	2.6	210	4	US-08-297-431B-34	Sequence 34, Appl	792	5	2.6	262	4	US-08-491-944-2	Sequence 2, Appl
720	5	2.6	210	4	US-09-443-948-4	Sequence 4, Appl	793	5	2.6	263	2	US-08-149-097D-28	Sequence 28, Appl
721	5	2.6	214	2	US-08-480-753-6	Sequence 6, Appl	794	5	2.6	263	4	US-08-818-112-91	Sequence 91, Appl
722	5	2.6	214	2	US-08-846-790A-1	Sequence 1, Appl	795	5	2.6	263	4	US-08-818-111-92	Sequence 92, Appl
723	5	2.6	214	3	US-08-935-333-1	Sequence 1, Appl	796	5	2.6	263	4	US-09-056-556-91	Sequence 91, Appl
724	5	2.6	214	3	US-09-041-889-11	Sequence 11, Appl	797	5	2.6	263	4	US-09-072-596-92	Sequence 92, Appl
725	5	2.6	214	3	US-08-837-058-11	Sequence 11, Appl	798	5	2.6	264	2	US-08-728-521-1	Sequence 1, Appl
726	5	2.6	215	2	US-08-531-525-49	Sequence 49, Appl	799	5	2.6	264	4	US-09-212-146-1	Sequence 1, Appl
727	5	2.6	215	2	US-08-718-270A-49	Sequence 49, Appl	800	5	2.6	264	4	US-09-134-001C-3780	Sequence 3780, Ap
728	5	2.6	215	4	US-09-240-410-4	Sequence 4, Appl	801	5	2.6	265	4	US-09-813-817-4	Sequence 4, Appl
729	5	2.6	218	1	US-07-607-538C-4	Sequence 4, Appl	802	5	2.6	265	4	US-09-134-001C-3336	Sequence 3336, Ap
730	5	2.6	218	2	US-08-162-402B-4	Sequence 4, Appl	803	5	2.6	265	4	US-09-978-197-4	Sequence 4, Appl
731	5	2.6	218	4	US-08-914-375C-6	Sequence 6, Appl	804	5	2.6	269	4	US-09-082-649B-82	Sequence 82, Appl
732	5	2.6	219	4	US-09-134-001C-5651	Sequence 5651, Ap	805	5	2.6	271	1	US-07-977-630-5	Sequence 5, Appl
733	5	2.6	221	2	US-08-861-269-6	Sequence 6, Appl	806	5	2.6	271	2	US-08-850-880-6	Sequence 6, Appl
734	5	2.6	221	2	US-09-134-596-6	Sequence 6, Appl	807	5	2.6	271	2	US-08-944-916-6	Sequence-6, Appl
735	5	2.6	221	3	US-09-293-273-6	Sequence 6, Appl	808	5	2.6	271	2	US-08-814-877-6	Sequence 6, Appl
736	5	2.6	223	4	US-09-230-196-17	Sequence 17, Appl	809	5	2.6	271	4	US-08-760-615-8	Sequence 8, Appl
737	5	2.6	223	4	US-09-230-196-24	Sequence 24, Appl	810	5	2.6	271	4	US-09-131-028A-5	Sequence 5, Appl
738	5	2.6	226	3	US-08-966-318-1	Sequence 1, Appl	811	5	2.6	271	4	US-09-131-028A-7	Sequence 7, Appl
739	5	2.6	226	4	US-09-216-619-1	Sequence 1, Appl	812	5	2.6	272	1	US-09-272-432A-6	Sequence 6, Appl
740	5	2.6	226	4	US-09-163-285-4	Sequence 4, Appl	813	5	2.6	272	1	US-08-252-995D-14	Sequence 14, Appl
741	5	2.6	226	4	US-09-134-001C-5520	Sequence 5520, Ap	814	5	2.6	273	2	US-08-834-108-14	Sequence 14, Appl
742	5	2.6	227	4	US-09-134-001C-4836	Sequence 4836, Ap	815	5	2.6	273	2	US-08-989-386-8	Sequence 8, Appl
743	5	2.6	228	2	US-08-408-095-21	Sequence 21, Appl	816	5	2.6	273	2	US-09-134-001C-3917	Sequence 3917, Ap
744	5	2.6	229	2	US-08-726-306A-23	Sequence 23, Appl	817	5	2.6	275	2	US-08-645-193B-19	Sequence 19, Appl
745	5	2.6	229	3	US-08-840-146-20	Sequence 20, Appl	818	5	2.6	276	1	US-07-952-817-27	Sequence 27, Appl
746	5	2.6	234	1	US-09-360-220-20	Sequence 20, Appl	819	5	2.6	278	2	US-08-685-992-34	Sequence 34, Appl
747	5	2.6	234	1	US-08-684-862-3	Sequence 3, Appl	820	5	2.6	278	2	US-09-144-925-34	Sequence 34, Appl
748	5	2.6	234	2	US-08-738-413B-12	Sequence 12, Appl	821	5	2.6	278	4	US-08-339-214-16	Sequence 16, Appl
749	5	2.6	234	4	US-09-134-001C-3195	Sequence 3195, Ap	822	5	2.6	278	4	US-08-339-214-26	Sequence 26, Appl
750	5	2.6	235	3	US-08-812-586-16	Sequence 16, Appl	823	5	2.6	278	6	5340934-8	Patent No. 5340934
751	5	2.6	235	4	US-09-029-785-1	Sequence 1, Appl	824	5	2.6	279	4	US-09-330-970-5	Sequence 5, Appl
752	5	2.6	238	4	US-09-370-838-37	Sequence 37, Appl	825	5	2.6	279	4	US-08-339-214-24	Sequence 24, Appl
753	5	2.6	239	2	US-08-286-819A-44	Sequence 44, Appl	826	5	2.6	279	5	US-08-339-214-32	Sequence 32, Appl
754	5	2.6	239	3	US-08-980-357-44	Sequence 44, Appl	827	5	2.6	279	5	PCT-US95-00362-5	Sequence 5, Appl
755	5	2.6	239	3	US-08-812-586-29	Sequence 29, Appl	828	5	2.6	281	2	US-08-810-453-2	Sequence 2, Appl
756	5	2.6	240	4	US-09-194-905-8	Sequence 8, Appl	829	5	2.6	281	3	US-08-815-190A-2	Sequence 2, Appl
757	5	2.6	241	4	US-09-387-372-3	Sequence 3, Appl	830	5	2.6	281	4	US-09-290-640-25	Sequence 25, Appl

539	2.6	123	4	US-09-156-316-10	Sequence 10, Appl	612	5	2.6	155	2	US-08-484-101B-12	Sequence 12, Appl
540	2.6	124	1	US-09-134-001C-5344	Sequence 5344, Ap	613	5	2.6	155	4	US-08-714-524D-12	Sequence 12, Appl
541	2.6	125	1	US-08-197-496A-2	Sequence 2, Appli	614	5	2.6	156	2	US-09-070-060-7	Sequence 7, Appli
542	2.6	125	1	US-08-275-370-2	Sequence 2, Appli	615	5	2.6	156	3	US-09-051-969A-3	Sequence 3, Appli
543	2.6	125	1	US-08-462-949-24	Sequence 24, Appl	616	5	2.6	156	3	US-09-051-969A-4	Sequence 4, Appli
544	2.6	125	1	US-08-367-968-2	Sequence 2, Appli	617	5	2.6	156	3	US-09-357-746-7	Sequence 7, Appli
545	2.6	125	1	US-08-023-764B-24	Sequence 24, Appl	618	5	2.6	156	4	US-09-615-192A-340	Sequence 340, App
546	2.6	125	1	US-08-665-484-2	Sequence 2, Appli	619	5	2.6	158	3	US-09-053-197A-19	Sequence 19, Appl
547	2.6	125	3	US-08-665-259-35	Sequence 35, Appl	620	5	2.6	158	3	US-08-946-329A-66	Sequence 66, Appl
548	2.6	125	3	US-08-762-500-35	Sequence 35, Appl	621	5	2.6	158	4	US-09-091-725-25	Sequence 25, Appl
549	2.6	126	2	US-08-341-843B-36	Sequence 36, Appl	622	5	2.6	158	4	US-09-085-761A-19	Sequence 19, Appl
550	2.6	126	2	US-08-427-497E-41	Sequence 41, Appl	623	5	2.6	159	2	US-08-829-110-1	Sequence 1, Appli
551	2.6	126	4	US-09-209-525-45	Sequence 45, Appl	624	5	2.6	159	2	US-08-162-402B-16	Sequence 16, Appl
552	2.6	127	2	US-08-476-176B-6	Sequence 6, Appli	625	5	2.6	160	4	US-08-858-207A-367	Sequence 367, App
553	2.6	127	2	US-08-476-176B-8	Sequence 8, Appli	626	5	2.6	160	4	US-09-370-838-205	Sequence 205, App
554	2.6	127	2	US-08-476-176B-10	Sequence 10, Appl	627	5	2.6	161	2	US-08-360-606B-33	Sequence 33, Appl
555	2.6	127	3	US-08-127-721A-6	Sequence 6, Appli	628	5	2.6	164	4	US-09-134-001C-4630	Sequence 4630, Ap
556	2.6	127	3	US-08-127-721A-8	Sequence 8, Appli	629	5	2.6	165	2	US-08-718-599-1	Sequence 1, Appli
557	2.6	127	3	US-08-127-721A-10	Sequence 10, Appl	630	5	2.6	165	2	US-08-587-680A-27	Sequence 27, Appl
558	2.6	127	3	US-08-485-246A-6	Sequence 6, Appli	631	5	2.6	166	3	US-08-765-381-4	Sequence 4, Appli
559	2.6	127	3	US-08-485-246A-8	Sequence 8, Appli	632	5	2.6	166	3	US-08-765-381-12	Sequence 12, Appl
560	2.6	127	3	US-08-485-246A-10	Sequence 10, Appl	633	5	2.6	166	4	US-09-134-001C-4279	Sequence 4279, Ap
561	2.6	128	2	US-08-557-309B-41	Sequence 41, Appl	634	5	2.6	168	1	US-08-145-995A-12	Sequence 12, Appl
562	2.6	128	3	US-08-834-306-41	Sequence 41, Appl	635	5	2.6	168	2	US-08-451-747-12	Sequence 12, Appl
563	2.6	128	4	US-08-993-674A-41	Sequence 41, Appl	636	5	2.6	168	3	US-08-946-329A-67	Sequence 67, Appl
564	2.6	128	4	US-09-256-976-41	Sequence 41, Appl	637	5	2.6	168	3	US-09-134-852-12	Sequence 12, Appl
565	2.6	130	4	US-08-339-214-28	Sequence 28, Appl	638	5	2.6	174	4	US-09-149-476-424	Sequence 424, App
566	2.6	134	1	US-08-405-034-4	Sequence 4, Appli	639	5	2.6	176	1	US-08-036-210-9	Sequence 9, Appli
567	2.6	136	4	US-08-339-214-83	Sequence 83, Appl	640	5	2.6	176	2	US-08-449-609-9	Sequence 9, Appli
568	2.6	137	3	US-08-685-871-60	Sequence 60, Appl	641	5	2.6	178	4	US-08-339-214-84	Sequence 84, Appl
569	2.6	137	4	US-08-339-214-20	Sequence 20, Appl	642	5	2.6	178	4	US-09-134-001C-3412	Sequence 3412, Ap
570	2.6	137	4	US-08-339-214-10	Sequence 10, Appl	643	5	2.6	179	3	US-08-649-100-9	Sequence 9, Appli
571	2.6	137	4	US-08-339-214-18	Sequence 18, Appl	644	5	2.6	179	4	US-08-874-102-41	Sequence 41, Appl
572	2.6	137	4	US-08-339-214-82	Sequence 82, Appl	645	5	2.6	179	4	US-08-874-102-44	Sequence 44, Appl
573	2.6	138	3	US-08-930-894-5	Sequence 5, Appli	646	5	2.6	179	4	US-08-339-214-6	Sequence 6, Appli
574	2.6	138	4	US-08-339-214-4	Sequence 4, Appli	647	5	2.6	179	4	US-08-339-214-14	Sequence 14, Appl
575	2.6	138	4	US-08-339-214-12	Sequence 12, Appl	648	5	2.6	179	4	US-08-339-214-22	Sequence 22, Appl
576	2.6	138	4	US-08-339-214-20	Sequence 20, Appl	649	5	2.6	179	4	US-08-984-919A-41	Sequence 41, Appl
577	2.6	138	4	US-08-339-214-81	Sequence 81, Appl	650	5	2.6	179	4	US-08-984-919A-44	Sequence 44, Appl
578	2.6	139	3	US-08-753-007A-33	Sequence 33, Appl	651	5	2.6	179	5	PCT-US93-04365-7	Sequence 7, Appli
579	2.6	139	4	US-09-398-496-33	Sequence 33, Appl	652	5	2.6	180	4	US-09-134-001C-2966	Sequence 2966, Ap
580	2.6	139	4	US-08-339-214-80	Sequence 80, Appl	653	5	2.6	181	3	US-08-753-007A-4	Sequence 4, Appli
581	2.6	139	4	US-08-336-165A-409	Sequence 409, App	654	5	2.6	181	4	US-09-398-496-4	Sequence 4, Appli
582	2.6	139	4	US-09-134-001C-5124	Sequence 5124, Ap	655	5	2.6	182	1	US-07-991-867B-10	Sequence 10, Appl
583	2.6	140	2	US-08-477-451-39	Sequence 39, Appl	656	5	2.6	182	1	US-08-117-083-10	Sequence 10, Appl
584	2.6	141	2	US-08-658-639-14	Sequence 14, Appl	657	5	2.6	182	1	US-08-107-755A-10	Sequence 10, Appl
585	2.6	141	4	US-08-944-604-14	Sequence 14, Appl	658	5	2.6	182	2	US-08-544-332-10	Sequence 10, Appl
586	2.6	141	4	US-09-286-529-22	Sequence 22, Appl	659	5	2.6	182	4	US-09-134-001C-3742	Sequence 3742, Ap
587	2.6	141	4	US-08-339-214-79	Sequence 79, Appl	660	5	2.6	182	4	US-09-370-861A-10	Sequence 10, Appl
588	2.6	142	4	US-09-134-001C-3473	Sequence 3473, Ap	661	5	2.6	183	3	US-09-122-443-11	Sequence 11, Appl
589	2.6	143	1	US-07-661-610C-8	Sequence 8, Appli	662	5	2.6	186	1	US-08-928-443-3	Sequence 3, Appli
590	2.6	143	4	US-09-325-932A-197	Sequence 197, App	663	5	2.6	186	3	US-09-129-055-3	Sequence 3, Appli
591	2.6	145	2	US-08-788-584-5	Sequence 5, Appli	664	5	2.6	187	2	US-08-177-109A-61	Sequence 61, Appl
592	2.6	145	3	US-08-946-914-12	Sequence 12, Appl	665	5	2.6	187	2	US-08-687-706-61	Sequence 61, Appl
593	2.6	145	3	US-08-630-172-5	Sequence 5, Appli	666	5	2.6	188	4	US-08-790-186A-1	Sequence 1, Appli
594	2.6	145	4	US-09-375-419-5	Sequence 5, Appli	667	5	2.6	188	4	US-08-861-269-7	Sequence 7, Appli
595	2.6	145	4	US-09-134-001C-5194	Sequence 5194, Ap	668	5	2.6	189	3	US-09-134-596-7	Sequence 7, Appli
596	2.6	145	4	US-09-656-450-12	Sequence 12, Appl	669	5	2.6	189	3	US-09-293-273-7	Sequence 7, Appli
597	2.6	146	4	US-09-209-525-2	Sequence 2, Appli	670	5	2.6	190	4	US-08-679-493A-184	Sequence 184, App
598	2.6	147	2	US-08-771-201-11	Sequence 11, Appl	671	5	2.6	191	2	US-08-290-665A-198	Sequence 198, App
599	2.6	147	3	US-08-946-329A-64	Sequence 64, Appl	672	5	2.6	191	2	US-08-290-665A-199	Sequence 199, App
600	2.6	147	4	US-09-134-001C-4542	Sequence 4542, Ap	673	5	2.6	191	2	US-08-290-665A-200	Sequence 200, App
601	2.6	147	4	US-09-134-001C-5518	Sequence 5518, Ap	674	5	2.6	191	2	US-08-290-665A-201	Sequence 201, App
602	2.6	148	3	US-08-946-329A-60	Sequence 60, Appl	675	5	2.6	191	2	US-08-290-665A-202	Sequence 202, App
603	2.6	148	3	US-08-946-329A-63	Sequence 63, Appl	676	5	2.6	191	2	US-08-290-665A-203	Sequence 203, App
604	2.6	148	3	US-08-946-329A-65	Sequence 65, Appl	677	5	2.6	191	2	US-08-290-665A-204	Sequence 204, App
605	2.6	148	4	US-09-352-990-4	Sequence 4, Appli	678	5	2.6	191	5	PCT-US95-10398-198	Sequence 198, App
606	2.6	148	4	US-09-171-461-18	Sequence 18, Appl	679	5	2.6	191	5	PCT-US95-10398-199	Sequence 199, App
607	2.6	149	3	US-08-584-031-17	Sequence 17, Appl	680	5	2.6	191	5	PCT-US95-10398-200	Sequence 200, App
608	2.6	150	4	US-08-862-124-5	Sequence 5, Appli	681	5	2.6	191	5	PCT-US95-10398-201	Sequence 201, App
609	2.6	150	4	US-09-134-001C-5014	Sequence 5014, Ap	682	5	2.6	191	5	PCT-US95-10398-202	Sequence 202, App
610	2.6	151	4	US-08-858-207A-298	Sequence 298, Appl	683	5	2.6	191	5	PCT-US95-10398-203	Sequence 203, App
611	2.6	155	1	US-08-530-010-12	Sequence 12, Appl	684	5	2.6	191	5	PCT-US95-10398-204	Sequence 204, App

393	5	2.6	93	4	US-09-134-001C-4887	Sequence 4887, Ap	466	5	2.6	108	1	US-08-276-852-99	Sequence 99, Appl
394	5	2.6	93	4	US-09-134-001C-4976	Sequence 4976, Ap	467	5	2.6	108	1	US-08-276-852-110	Sequence 110, Appl
395	5	2.6	93	5	PCT-US95-06266-53	Sequence 53, Appl	468	5	2.6	108	1	US-08-474-040-86	Sequence 86, Appl
396	5	2.6	94	5	PCT-US95-08743-111	Sequence 111, Appl	469	5	2.6	108	1	US-08-487-200-86	Sequence 86, Appl
397	5	2.6	94	1	US-08-466-033-67	Sequence 67, Appl	470	5	2.6	108	1	US-08-488-113B-150	Sequence 150, Appl
398	5	2.6	94	2	US-08-444-733-67	Sequence 67, Appl	471	5	2.6	108	1	US-08-477-484B-150	Sequence 150, Appl
399	5	2.6	94	2	US-08-464-134-67	Sequence 67, Appl	472	5	2.6	108	1	US-08-899-575-86	Sequence 86, Appl
400	5	2.6	94	2	US-08-461-361-67	Sequence 67, Appl	473	5	2.6	108	1	US-08-899-575-88	Sequence 88, Appl
401	5	2.6	94	2	US-08-485-910-67	Sequence 67, Appl	474	5	2.6	108	1	US-08-899-575-96	Sequence 96, Appl
402	5	2.6	94	5	PCT-US95-06266-51	Sequence 51, Appl	475	5	2.6	108	1	US-08-899-575-99	Sequence 99, Appl
403	5	2.6	97	2	US-09-109-266-20	Sequence 20, Appl	476	5	2.6	108	1	US-08-899-575-110	Sequence 110, Appl
404	5	2.6	98	4	US-09-183-861-42	Sequence 42, Appl	477	5	2.6	108	1	US-08-899-575-86	Sequence 86, Appl
405	5	2.6	98	4	US-09-022-765-42	Sequence 42, Appl	478	5	2.6	108	1	US-08-899-575-88	Sequence 88, Appl
406	5	2.6	101	4	US-09-482-611B-99	Sequence 99, Appl	479	5	2.6	108	1	US-08-899-575-96	Sequence 96, Appl
407	5	2.6	101	4	US-09-482-611B-101	Sequence 101, Appl	480	5	2.6	108	1	US-08-899-575-99	Sequence 99, Appl
408	5	2.6	103	2	US-08-771-201-9	Sequence 9, Appl	481	5	2.6	108	1	US-08-899-575-110	Sequence 110, Appl
409	5	2.6	104	1	US-08-276-852-92	Sequence 92, Appl	482	5	2.6	108	2	US-08-646-360-150	Sequence 150, Appl
410	5	2.6	104	1	US-08-276-852-100	Sequence 100, Appl	483	5	2.6	108	2	US-08-232-081B-42	Sequence 42, Appl
411	5	2.6	104	1	US-08-899-575-92	Sequence 92, Appl	484	5	2.6	108	4	US-08-839-765-150	Sequence 150, Appl
412	5	2.6	104	1	US-08-899-575-100	Sequence 100, Appl	485	5	2.6	108	4	US-09-136-389-150	Sequence 150, Appl
413	5	2.6	104	1	US-08-899-575-92	Sequence 92, Appl	486	5	2.6	108	4	US-08-484-537-86	Sequence 86, Appl
414	5	2.6	104	1	US-08-899-575-100	Sequence 100, Appl	487	5	2.6	108	4	US-09-240-274-178	Sequence 178, Appl
415	5	2.6	104	5	PCT-US95-08743-92	Sequence 92, Appl	488	5	2.6	108	4	US-09-610-838-150	Sequence 150, Appl
416	5	2.6	104	5	PCT-US95-08743-100	Sequence 100, Appl	489	5	2.6	108	4	US-09-134-001C-4536	Sequence 4536, Ap
417	5	2.6	105	4	US-08-857-076-72	Sequence 72, Appl	490	5	2.6	108	5	PCT-US95-08743-86	Sequence 86, Appl
418	5	2.6	105	4	US-09-134-001C-1949	Sequence 4949, Ap	491	5	2.6	108	5	PCT-US95-08743-88	Sequence 88, Appl
419	5	2.6	107	1	US-07-634-278-63	Sequence 63, Appl	492	5	2.6	108	5	PCT-US95-08743-96	Sequence 96, Appl
420	5	2.6	107	1	US-07-634-278-87	Sequence 87, Appl	493	5	2.6	108	5	PCT-US95-08743-99	Sequence 99, Appl
421	5	2.6	107	1	US-08-477-728-63	Sequence 63, Appl	494	5	2.6	108	5	PCT-US95-08743-110	Sequence 110, Appl
422	5	2.6	107	1	US-08-477-728-87	Sequence 87, Appl	495	5	2.6	109	1	US-08-276-852-147	Sequence 147, Appl
423	5	2.6	107	1	US-08-036-210-4	Sequence 4, Appl	496	5	2.6	109	1	US-08-162-102C-23	Sequence 23, Appl
424	5	2.6	107	1	US-08-276-852-90	Sequence 90, Appl	497	5	2.6	109	1	US-08-899-575-147	Sequence 147, Appl
425	5	2.6	107	1	US-08-276-852-116	Sequence 116, Appl	498	5	2.6	109	1	US-08-899-575-149	Sequence 149, Appl
426	5	2.6	107	1	US-08-276-852-117	Sequence 117, Appl	499	5	2.6	109	4	US-09-025-769B-16	Sequence 16, Appl
427	5	2.6	107	1	US-08-276-852-118	Sequence 118, Appl	500	5	2.6	109	5	PCT-US93-08786-23	Sequence 23, Appl
428	5	2.6	107	1	US-08-276-852-121	Sequence 121, Appl	501	5	2.6	109	5	PCT-US95-08743-147	Sequence 147, Appl
429	5	2.6	107	1	US-08-276-852-122	Sequence 122, Appl	502	5	2.6	110	4	US-09-209-525-52	Sequence 52, Appl
430	5	2.6	107	1	US-08-474-040-63	Sequence 63, Appl	503	5	2.6	110	4	US-08-936-165A-391	Sequence 391, Appl
431	5	2.6	107	1	US-08-474-040-87	Sequence 87, Appl	504	5	2.6	111	1	US-07-634-278-47	Sequence 47, Appl
432	5	2.6	107	1	US-08-487-200-63	Sequence 63, Appl	505	5	2.6	111	1	US-08-477-728-47	Sequence 47, Appl
433	5	2.6	107	1	US-08-487-200-87	Sequence 87, Appl	506	5	2.6	111	1	US-08-276-852-149	Sequence 149, Appl
434	5	2.6	107	1	US-08-107-669D-14	Sequence 14, Appl	507	5	2.6	111	1	US-08-474-040-47	Sequence 47, Appl
435	5	2.6	107	1	US-08-472-788A-14	Sequence 14, Appl	508	5	2.6	111	1	US-08-487-200-47	Sequence 47, Appl
436	5	2.6	107	1	US-08-899-575-90	Sequence 90, Appl	509	5	2.6	111	1	US-08-899-575-149	Sequence 149, Appl
437	5	2.6	107	1	US-08-899-575-116	Sequence 116, Appl	510	5	2.6	111	1	US-08-899-575-149	Sequence 149, Appl
438	5	2.6	107	1	US-08-899-575-117	Sequence 117, Appl	511	5	2.6	111	4	US-08-484-537-47	Sequence 47, Appl
439	5	2.6	107	1	US-08-899-575-118	Sequence 118, Appl	512	5	2.6	111	5	PCT-US95-08743-149	Sequence 149, Appl
440	5	2.6	107	1	US-08-899-575-121	Sequence 121, Appl	513	5	2.6	112	1	US-08-276-852-151	Sequence 151, Appl
441	5	2.6	107	1	US-08-899-575-122	Sequence 122, Appl	514	5	2.6	112	1	US-08-899-575-151	Sequence 151, Appl
442	5	2.6	107	1	US-08-899-575-90	Sequence 90, Appl	515	5	2.6	112	1	US-08-899-575-151	Sequence 151, Appl
443	5	2.6	107	1	US-08-899-575-116	Sequence 116, Appl	516	5	2.6	112	2	US-08-475-000-18	Sequence 18, Appl
444	5	2.6	107	1	US-08-899-575-117	Sequence 117, Appl	517	5	2.6	112	2	US-08-483-199-18	Sequence 18, Appl
445	5	2.6	107	1	US-08-899-575-118	Sequence 118, Appl	518	5	2.6	112	2	US-08-484-508-18	Sequence 18, Appl
446	5	2.6	107	1	US-08-899-575-121	Sequence 121, Appl	519	5	2.6	112	5	PCT-US95-08743-151	Sequence 151, Appl
447	5	2.6	107	2	US-08-899-575-122	Sequence 122, Appl	520	5	2.6	114	2	US-08-771-201-10	Sequence 10, Appl
448	5	2.6	107	2	US-08-477-531B-14	Sequence 14, Appl	521	5	2.6	115	1	US-08-152-922A-7	Sequence 7, Appl
449	5	2.6	107	2	US-08-652-558-49	Sequence 49, Appl	522	5	2.6	115	1	US-08-111-939-22	Sequence 22, Appl
450	5	2.6	107	2	US-08-082-842A-14	Sequence 14, Appl	523	5	2.6	116	1	US-08-053-131-183	Sequence 183, Appl
451	5	2.6	107	2	US-08-232-081B-9	Sequence 9, Appl	524	5	2.6	116	2	US-08-096-762-183	Sequence 183, Appl
452	5	2.6	107	2	US-08-449-609-4	Sequence 4, Appl	525	5	2.6	116	4	US-09-042-353-46	Sequence 46, Appl
453	5	2.6	107	4	US-08-484-537-63	Sequence 63, Appl	526	5	2.6	117	4	US-08-758-417A-311	Sequence 311, Appl
454	5	2.6	107	4	US-08-484-537-87	Sequence 87, Appl	527	5	2.6	117	4	US-09-071-035-354	Sequence 354, Appl
455	5	2.6	107	5	PCT-US95-08743-90	Sequence 90, Appl	528	5	2.6	119	2	US-08-652-816A-10	Sequence 10, Appl
456	5	2.6	107	5	PCT-US95-08743-116	Sequence 116, Appl	529	5	2.6	121	1	US-08-466-033-51	Sequence 51, Appl
457	5	2.6	107	5	PCT-US95-08743-117	Sequence 117, Appl	530	5	2.6	121	2	US-08-444-733-51	Sequence 51, Appl
458	5	2.6	107	5	PCT-US95-08743-118	Sequence 118, Appl	531	5	2.6	121	2	US-08-464-134-51	Sequence 51, Appl
459	5	2.6	107	5	PCT-US95-08743-121	Sequence 121, Appl	532	5	2.6	121	2	US-08-461-361-51	Sequence 51, Appl
460	5	2.6	107	5	PCT-US95-08743-122	Sequence 122, Appl	533	5	2.6	121	2	US-08-485-910-51	Sequence 51, Appl
461	5	2.6	108	1	US-07-634-278-86	Sequence 86, Appl	534	5	2.6	121	2	US-08-658-639-13	Sequence 13, Appl
462	5	2.6	108	1	US-08-477-728-86	Sequence 86, Appl	535	5	2.6	121	4	US-08-944-604-13	Sequence 13, Appl
463	5	2.6	108	1	US-08-276-852-86	Sequence 86, Appl	536	5	2.6	121	4	US-09-591-435-13	Sequence 13, Appl
464	5	2.6	108	1	US-08-276-852-88	Sequence 88, Appl	537	5	2.6	121	4	US-09-342-647-12	Sequence 12, Appl
465	5	2.6	108	1	US-08-276-852-96	Sequence 96, Appl	538	5	2.6	121	5	PCT-US95-06266-35	Sequence 35, Appl

247	5	2.6	30	1	US-08-056-200-16	Sequence 16, Appl	320	5	2.6	68	4	US-08-858-207A-546	Sequence 546, App
248	5	2.6	30	2	US-08-800-644-16	Sequence 16, Appl	321	5	2.6	68	4	US-09-527-236A-22	Sequence 22, Appl
249	5	2.6	32	1	US-08-190-802A-184	Sequence 184, App	322	5	2.6	69	4	US-08-358-160-125	Sequence 125, App
250	5	2.6	32	4	US-08-477-346-184	Sequence 184, App	323	5	2.6	70	1	US-08-330-163-27	Sequence 27, Appl
251	5	2.6	32	4	US-08-473-089-184	Sequence 184, App	324	5	2.6	70	1	US-08-482-111-27	Sequence 27, Appl
252	5	2.6	32	4	US-08-487-072A-184	Sequence 184, App	325	5	2.6	70	1	US-09-620-093A-6	Sequence 6, Appl
253	5	2.6	34	2	US-08-557-309B-66	Sequence 66, Appl	326	5	2.6	73	4	US-09-149-476-579	Sequence 579, App
254	5	2.6	35	1	US-08-151-574-2	Sequence 2, Appl	327	5	2.6	74	4	US-09-209-525-51	Sequence 51, Appl
255	5	2.6	35	2	US-08-419-448-2	Sequence 2, Appl	328	5	2.6	75	1	US-08-350-884-35	Sequence 35, Appl
256	5	2.6	35	4	US-08-974-549A-53	Sequence 53, Appl	329	5	2.6	75	1	US-08-709-173-35	Sequence 35, Appl
257	5	2.6	35	4	US-09-233-510-2	Sequence 2, Appl	330	5	2.6	75	2	US-08-709-173-35	Sequence 35, Appl
258	5	2.6	36	3	US-09-045-764A-12	Sequence 12, Appl	331	5	2.6	75	2	US-08-833-678A-3	Sequence 3, Appl
259	5	2.6	40	2	US-08-288-771-7	Sequence 7, Appl	332	5	2.6	75	4	US-08-529-169A-3	Sequence 3, Appl
260	5	2.6	40	4	US-09-162-777-7	Sequence 7, Appl	333	5	2.6	76	1	US-08-232-815-2	Sequence 2, Appl
261	5	2.6	41	2	US-08-288-771-6	Sequence 6, Appl	334	5	2.6	76	1	US-08-350-906-2	Sequence 2, Appl
262	5	2.6	41	4	US-09-162-777-6	Sequence 6, Appl	335	5	2.6	76	1	US-08-551-275-2	Sequence 2, Appl
263	5	2.6	42	2	US-08-288-771-5	Sequence 5, Appl	336	5	2.6	76	1	US-08-450-834-6	Sequence 6, Appl
264	5	2.6	42	4	US-08-751-344B-4	Sequence 4, Appl	337	5	2.6	76	2	US-08-743-200-14	Sequence 14, Appl
265	5	2.6	42	4	US-09-162-777-5	Sequence 5, Appl	338	5	2.6	76	2	US-09-070-060-8	Sequence 8, Appl
266	5	2.6	44	1	US-08-363-311-29	Sequence 29, Appl	339	5	2.6	76	3	US-09-357-746-8	Sequence 8, Appl
267	5	2.6	44	2	US-08-463-288A-29	Sequence 29, Appl	340	5	2.6	76	4	US-08-817-787-1	Sequence 1, Appl
268	5	2.6	44	2	US-08-470-445A-29	Sequence 29, Appl	341	5	2.6	76	4	US-08-817-787-8	Sequence 8, Appl
269	5	2.6	44	2	US-08-462-679-29	Sequence 29, Appl	342	5	2.6	76	4	US-09-100-802-1	Sequence 1, Appl
270	5	2.6	44	2	US-08-466-210A-29	Sequence 29, Appl	343	5	2.6	76	5	US-09-331-930A-26	Sequence 26, Appl
271	5	2.6	44	2	US-08-467-147A-29	Sequence 29, Appl	344	5	2.6	76	5	PCT-US95-04536-2	Sequence 2, Appl
272	5	2.6	44	2	US-08-469-014-29	Sequence 29, Appl	345	5	2.6	76	5	PCT-US95-13658-6	Sequence 6, Appl
273	5	2.6	44	4	US-09-046-894-41	Sequence 41, Appl	346	5	2.6	77	4	US-09-227-357-552	Sequence 552, App
274	5	2.6	44	4	US-09-046-894-42	Sequence 42, Appl	347	5	2.6	77	4	US-08-828-683A-24	Sequence 24, Appl
275	5	2.6	44	4	US-08-687-590-13	Sequence 13, Appl	348	5	2.6	77	6	5510474-3	Patent No. 5510474
276	5	2.6	44	4	US-08-687-590-14	Sequence 14, Appl	349	5	2.6	78	2	US-08-117-952-749	Sequence 749, App
277	5	2.6	44	4	US-08-687-590-15	Sequence 15, Appl	350	5	2.6	78	3	US-08-505-486-94	Sequence 94, Appl
278	5	2.6	44	4	US-09-346-290-29	Sequence 29, Appl	351	5	2.6	78	3	US-08-801-028-94	Sequence 94, Appl
279	5	2.6	44	5	PCT-US93-10506A-29	Sequence 29, Appl	352	5	2.6	78	3	US-09-340-154-94	Sequence 94, Appl
280	5	2.6	44	5	PCT-US93-10508-29	Sequence 29, Appl	353	5	2.6	78	4	US-09-482-611B-94	Sequence 94, Appl
281	5	2.6	45	2	US-08-219-237B-10	Sequence 10, Appl	354	5	2.6	78	4	US-09-482-611B-100	Sequence 100, App
282	5	2.6	45	4	US-09-209-525-50	Sequence 50, Appl	355	5	2.6	78	4	US-09-482-611B-102	Sequence 102, App
283	5	2.6	47	4	US-09-314-268-154	Sequence 154, App	356	5	2.6	78	5	PCT-US95-09338-94	Sequence 94, Appl
284	5	2.6	47	4	US-09-463-238-9	Sequence 9, Appl	357	5	2.6	78	5	PCT-US95-09339-94	Sequence 94, Appl
285	5	2.6	48	1	US-08-290-919-1	Sequence 1, Appl	358	5	2.6	81	2	US-09-014-969-32	Sequence 32, Appl
286	5	2.6	48	1	US-08-290-919-2	Sequence 2, Appl	359	5	2.6	82	1	US-08-476-405A-25	Sequence 25, Appl
287	5	2.6	48	1	US-08-290-919-12	Sequence 12, Appl	360	5	2.6	83	2	US-08-248-839C-97	Sequence 97, Appl
288	5	2.6	48	1	US-08-602-010A-26	Sequence 26, Appl	361	5	2.6	84	4	US-09-406-045-6	Sequence 6, Appl
289	5	2.6	48	1	US-08-466-033-272	Sequence 272, App	362	5	2.6	85	1	US-08-615-170-22	Sequence 22, Appl
290	5	2.6	48	1	US-08-680-726A-26	Sequence 26, Appl	363	5	2.6	85	4	US-09-042-785A-26	Sequence 26, Appl
291	5	2.6	48	2	US-08-444-733-272	Sequence 272, App	364	5	2.6	86	4	US-09-208-277-5	Sequence 5, Appl
292	5	2.6	48	2	US-08-464-134-272	Sequence 272, App	365	5	2.6	86	4	US-09-556-877-5	Sequence 5, Appl
293	5	2.6	48	2	US-08-461-361-272	Sequence 272, App	366	5	2.6	86	4	US-09-288-594A-5	Sequence 5, Appl
294	5	2.6	48	2	US-08-485-910-272	Sequence 272, App	367	5	2.6	86	4	US-09-620-412C-5	Sequence 5, Appl
295	5	2.6	48	4	US-09-092-409-26	Sequence 26, Appl	368	5	2.6	87	3	US-08-906-769-161	Sequence 161, App
296	5	2.6	52	1	US-08-162-102C-43	Sequence 43, Appl	369	5	2.6	87	3	US-08-906-616-161	Sequence 161, App
297	5	2.6	52	3	US-08-908-643C-61	Sequence 61, Appl	370	5	2.6	87	4	US-08-639-075A-161	Sequence 161, App
298	5	2.6	58	1	US-08-055-988-1	Sequence 1, Appl	371	5	2.6	87	4	US-09-012-431-161	Sequence 161, App
299	5	2.6	60	4	US-08-751-344B-5	Sequence 5, Appl	372	5	2.6	87	4	US-09-012-692-161	Sequence 161, App
300	5	2.6	61	2	US-08-757-316C-30	Sequence 30, Appl	373	5	2.6	87	4	US-08-906-613-161	Sequence 161, App
301	5	2.6	61	4	US-09-208-277-6	Sequence 6, Appl	374	5	2.6	88	4	US-09-330-970-6	Sequence 6, Appl
302	5	2.6	61	4	US-09-416-050A-27	Sequence 27, Appl	375	5	2.6	88	3	US-08-906-769-165	Sequence 165, App
303	5	2.6	61	4	US-09-664-800-27	Sequence 27, Appl	376	5	2.6	89	3	US-08-906-616-165	Sequence 165, App
304	5	2.6	61	4	US-09-665-309-27	Sequence 27, Appl	377	5	2.6	89	4	US-08-639-075A-165	Sequence 165, App
305	5	2.6	61	4	US-09-661-569-27	Sequence 27, Appl	378	5	2.6	89	4	US-09-012-431-165	Sequence 165, App
306	5	2.6	61	4	US-09-556-877-6	Sequence 6, Appl	379	5	2.6	89	4	US-09-012-692-165	Sequence 165, App
307	5	2.6	61	4	US-09-288-594A-6	Sequence 6, Appl	380	5	2.6	89	4	US-08-906-613-165	Sequence 165, App
308	5	2.6	61	4	US-09-620-412C-6	Sequence 6, Appl	381	5	2.6	89	4	US-09-134-001C-5434	Sequence 5434, Ap
309	5	2.6	62	1	US-08-466-033-65	Sequence 65, Appl	382	5	2.6	91	4	US-09-071-035-356	Sequence 356, App
310	5	2.6	62	2	US-08-444-733-65	Sequence 65, Appl	383	5	2.6	92	1	US-08-181-492B-27	Sequence 27, Appl
311	5	2.6	62	2	US-08-464-134-65	Sequence 65, Appl	384	5	2.6	92	5	PCT-US95-00408-27	Sequence 27, Appl
312	5	2.6	62	2	US-08-461-361-65	Sequence 65, Appl	385	5	2.6	93	1	US-08-276-852-111	Sequence 111, App
313	5	2.6	62	2	US-08-485-910-65	Sequence 65, Appl	386	5	2.6	93	1	US-08-466-033-69	Sequence 69, Appl
314	5	2.6	62	5	PCT-US95-06266-49	Sequence 49, Appl	387	5	2.6	93	1	US-08-899-575-111	Sequence 111, App
315	5	2.6	63	4	US-08-894-626-1	Sequence 1, Appl	388	5	2.6	93	2	US-08-899-575-111	Sequence 111, App
316	5	2.6	65	1	US-08-227-536-4	Sequence 4, Appl	389	5	2.6	93	2	US-08-444-733-69	Sequence 69, Appl
317	5	2.6	65	5	PCT-US95-04682-4	Sequence 4, Appl	390	5	2.6	93	2	US-08-464-134-69	Sequence 69, Appl
318	5	2.6	66	4	US-09-134-001C-3495	Sequence 3495, Ap	391	5	2.6	93	2	US-08-461-361-69	Sequence 69, Appl
319	5	2.6	68	4	US-08-803-346-58	Sequence 58, Appl	392	5	2.6	93	2	US-08-485-910-69	Sequence 69, Appl

101	6	3.1	650	4	US-09-232-201-79	Sequence 79, Appl	174	5	2.6	12	2	US-08-476-062A-18	Sequence 18, Appl
102	6	3.1	650	4	US-09-232-201-97	Sequence 97, Appl	175	5	2.6	12	2	US-08-934-222-102	Sequence 102, App
103	6	3.1	727	2	US-08-475-844-9	Sequence 9, Appl	176	5	2.6	12	2	US-08-933-402-102	Sequence 102, App
104	6	3.1	727	5	PCT-US95-08429-9	Sequence 9, Appl	177	5	2.6	12	2	US-09-207-621-102	Sequence 102, App
105	6	3.1	728	2	US-08-475-844-5	Sequence 5, Appl	178	5	2.6	12	2	US-08-532-818-102	Sequence 102, App
106	6	3.1	728	5	PCT-US95-08429-5	Sequence 5, Appl	179	5	2.6	12	3	US-09-231-797-102	Sequence 102, App
107	6	3.1	752	2	US-08-896-590A-2	Sequence 2, Appl	180	5	2.6	12	3	US-08-934-224-102	Sequence 102, App
108	6	3.1	752	2	US-08-896-590A-4	Sequence 4, Appl	181	5	2.6	12	3	US-08-933-843-102	Sequence 102, App
109	6	3.1	752	4	US-09-185-832-2	Sequence 2, Appl	182	5	2.6	12	4	US-08-934-223-102	Sequence 102, App
110	6	3.1	752	4	US-09-185-832-4	Sequence 4, Appl	183	5	2.6	12	4	US-09-035-249A-18	Sequence 18, Appl
111	6	3.1	829	1	US-08-346-455B-34	Sequence 34, Appl	184	5	2.6	12	4	US-09-413-492-102	Sequence 102, App
112	6	3.1	829	3	US-08-977-221-34	Sequence 34, Appl	185	5	2.6	12	4	US-09-057-897-14	Sequence 14, Appl
113	6	3.1	829	4	US-09-483-831B-34	Sequence 34, Appl	186	5	2.6	12	5	PCT-US96-01314-18	Sequence 18, Appl
114	6	3.1	829	5	PCT-US95-06613-34	Sequence 34, Appl	187	5	2.6	13	1	US-08-452-722-1	Sequence 1, Appl
115	6	3.1	831	1	US-09-269-861A-8	Sequence 8, Appl	188	5	2.6	13	1	US-08-404-731A-1	Sequence 1, Appl
116	6	3.1	868	2	US-08-907-166-2	Sequence 2, Appl	189	5	2.6	13	2	US-08-344-227-1	Sequence 1, Appl
117	6	3.1	892	1	US-07-977-434-12	Sequence 12, Appl	190	5	2.6	13	2	US-08-503-226B-1	Sequence 1, Appl
118	6	3.1	892	1	US-08-458-819-12	Sequence 12, Appl	191	5	2.6	13	3	US-08-721-458B-1	Sequence 1, Appl
119	6	3.1	908	5	PCT-US91-07035-12	Sequence 12, Appl	192	5	2.6	14	4	US-09-112-096-11	Sequence 11, Appl
120	6	3.1	908	1	US-08-356-354-6	Sequence 6, Appl	193	5	2.6	15	1	US-08-452-722-2	Sequence 2, Appl
121	6	3.1	908	2	US-08-778-656-6	Sequence 6, Appl	194	5	2.6	15	1	US-08-452-722-3	Sequence 3, Appl
122	6	3.1	915	1	US-08-346-455B-69	Sequence 69, Appl	195	5	2.6	15	1	US-08-404-731A-2	Sequence 2, Appl
123	6	3.1	915	3	US-08-977-221-69	Sequence 69, Appl	196	5	2.6	15	1	US-08-404-731A-3	Sequence 3, Appl
124	6	3.1	915	4	US-09-483-831B-69	Sequence 69, Appl	197	5	2.6	15	1	US-08-476-405A-18	Sequence 18, Appl
125	6	3.1	1013	5	PCT-US95-06613-69	Sequence 69, Appl	198	5	2.6	15	1	US-08-476-405A-19	Sequence 19, Appl
126	6	3.1	1013	4	US-09-415-522-8	Sequence 8, Appl	199	5	2.6	15	1	US-08-344-227-2	Sequence 2, Appl
127	6	3.1	1054	1	US-08-356-354-4	Sequence 4, Appl	200	5	2.6	15	1	US-08-344-227-3	Sequence 3, Appl
128	6	3.1	1054	2	US-08-778-656-4	Sequence 4, Appl	201	5	2.6	15	2	US-08-480-190-240	Sequence 240, App
129	6	3.1	1119	4	US-09-396-651B-2	Sequence 2, Appl	202	5	2.6	15	2	US-08-503-226B-2	Sequence 2, Appl
130	6	3.1	1187	1	US-08-320-559-28	Sequence 28, Appl	203	5	2.6	15	2	US-08-503-226B-3	Sequence 3, Appl
131	6	3.1	1187	3	US-08-545-860D-28	Sequence 28, Appl	204	5	2.6	15	2	US-08-488-379-240	Sequence 240, App
132	6	3.1	1187	5	PCT-US94-04496-28	Sequence 28, Appl	205	5	2.6	15	3	US-08-721-458B-2	Sequence 2, Appl
133	6	3.1	1210	1	US-08-320-559-26	Sequence 26, Appl	206	5	2.6	15	3	US-08-721-458B-3	Sequence 3, Appl
134	6	3.1	1210	5	PCT-US94-04496-26	Sequence 26, Appl	207	5	2.6	15	4	US-09-057-897-6	Sequence 6, Appl
135	6	3.1	1210	1	US-08-545-860D-26	Sequence 26, Appl	208	5	2.6	15	4	US-09-209-525-46	Sequence 46, Appl
136	6	3.1	1338	2	US-08-471-033-50	Sequence 50, Appl	209	5	2.6	15	4	US-09-556-877-100	Sequence 100, App
137	6	3.1	1338	2	US-08-471-044-50	Sequence 50, Appl	210	5	2.6	15	4	US-09-620-412C-100	Sequence 100, App
138	6	3.1	1338	2	US-08-463-483A-50	Sequence 50, Appl	211	5	2.6	15	5	PCT-US93-07545-240	Sequence 240, App
139	6	3.1	1338	2	US-08-471-046A-50	Sequence 50, Appl	212	5	2.6	16	4	US-09-057-897-12	Sequence 12, Appl
140	6	3.1	1338	2	US-08-470-566B-50	Sequence 50, Appl	213	5	2.6	17	4	US-09-057-052-19	Sequence 19, Appl
141	6	3.1	1338	2	US-08-469-334-50	Sequence 50, Appl	214	5	2.6	17	4	US-09-342-681C-122	Sequence 122, App
142	6	3.1	1338	3	US-09-300-529-50	Sequence 50, Appl	215	5	2.6	18	2	US-08-480-190-239	Sequence 239, App
143	6	3.1	1346	1	US-08-471-033-23	Sequence 23, Appl	216	5	2.6	18	2	US-08-488-379-239	Sequence 239, App
144	6	3.1	1346	2	US-08-471-044-23	Sequence 23, Appl	217	5	2.6	18	4	US-09-057-897-5	Sequence 5, Appl
145	6	3.1	1346	2	US-08-463-483A-23	Sequence 23, Appl	218	5	2.6	18	5	PCT-US93-07545-239	Sequence 239, App
146	6	3.1	1346	2	US-08-471-046A-23	Sequence 23, Appl	219	5	2.6	19	2	US-08-793-490-5	Sequence 5, Appl
147	6	3.1	1346	2	US-08-470-566B-23	Sequence 23, Appl	220	5	2.6	19	4	US-09-057-897-11	Sequence 11, Appl
148	6	3.1	1346	2	US-08-469-334-23	Sequence 23, Appl	221	5	2.6	19	4	US-09-310-187A-2	Sequence 2, Appl
149	6	3.1	1346	3	US-09-300-529-23	Sequence 23, Appl	222	5	2.6	19	4	US-08-469-260A-398	Sequence 398, App
150	6	3.1	1477	3	US-08-492-459-10	Sequence 10, Appl	223	5	2.6	19	4	US-09-441-502B-72	Sequence 72, Appl
151	6	3.1	1477	3	US-08-423-752-10	Sequence 10, Appl	224	5	2.6	20	1	US-08-476-405A-2	Sequence 2, Appl
152	6	3.1	1477	3	US-08-945-994-3	Sequence 3, Appl	225	5	2.6	20	5	PCT-US96-01314-65	Sequence 65, Appl
153	6	3.1	1477	4	US-08-716-873-24	Sequence 24, Appl	226	5	2.6	21	1	US-08-476-405A-3	Sequence 3, Appl
154	6	3.1	1477	4	US-09-368-431-24	Sequence 24, Appl	227	5	2.6	21	1	US-08-476-405A-4	Sequence 4, Appl
155	6	3.1	1477	4	US-09-414-006-10	Sequence 10, Appl	228	5	2.6	21	1	US-08-476-405A-13	Sequence 13, Appl
156	6	3.1	1477	4	US-09-447-223-10	Sequence 10, Appl	229	5	2.6	21	4	US-09-057-897-1	Sequence 1, Appl
157	6	3.1	1720	2	US-08-477-451-12	Sequence 12, Appl	230	5	2.6	21	4	US-09-057-897-32	Sequence 32, Appl
158	6	3.1	1850	4	US-09-620-093A-5	Sequence 5, Appl	231	5	2.6	22	4	US-09-248-588-24	Sequence 24, Appl
159	6	3.1	1898	1	US-08-056-200-94	Sequence 94, Appl	232	5	2.6	22	4	US-09-057-897-2	Sequence 2, Appl
160	6	3.1	1898	2	US-08-800-644-94	Sequence 94, Appl	233	5	2.6	22	2	US-08-288-771-4	Sequence 4, Appl
161	6	3.1	2052	3	US-09-045-201A-2	Sequence 2, Appl	234	5	2.6	23	4	US-09-162-777-4	Sequence 4, Appl
162	6	3.1	2052	4	US-09-619-062-2	Sequence 2, Appl	235	5	2.6	24	2	US-08-288-771-3	Sequence 3, Appl
163	5	2.6	6	1	US-08-290-919-14	Sequence 14, Appl	236	5	2.6	24	4	US-09-162-777-3	Sequence 3, Appl
164	5	2.6	7	4	US-09-479-431A-1	Sequence 1, Appl	237	5	2.6	25	1	US-08-444-005-21	Sequence 21, Appl
165	5	2.6	10	2	US-08-658-639-8	Sequence 8, Appl	238	5	2.6	25	4	US-09-578-303-28	Sequence 28, Appl
166	5	2.6	10	4	US-09-306-593-3	Sequence 3, Appl	239	5	2.6	25	4	US-08-469-260A-288	Sequence 288, App
167	5	2.6	10	4	US-08-944-604-8	Sequence 8, Appl	240	5	2.6	26	1	US-08-056-200-17	Sequence 17, Appl
168	5	2.6	10	4	US-09-295-996B-1	Sequence 1, Appl	241	5	2.6	26	1	US-08-056-200-18	Sequence 18, Appl
169	5	2.6	11	3	US-08-649-100-4	Sequence 4, Appl	242	5	2.6	26	2	US-07-942-245-483	Sequence 483, App
170	5	2.6	11	4	US-09-057-897-8	Sequence 8, Appl	243	5	2.6	26	2	US-08-800-644-17	Sequence 17, Appl
171	5	2.6	11	4	US-09-057-897-17	Sequence 17, Appl	244	5	2.6	26	2	US-08-800-644-18	Sequence 18, Appl
172	5	2.6	11	5	PCT-US91-08328-13	Sequence 13, Appl	245	5	2.6	27	4	US-08-974-549A-167	Sequence 167, App
173	5	2.6	12	1	US-08-156-552A-18	Sequence 18, Appl	246	5	2.6	28	2	US-08-620-151-4	Sequence 4, Appl

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2003, 09:59:03 ; Search time 15 seconds
(without alignments)
376.613 Million cell updates/sec

Title: US-09-868-352-23
Perfect score: 192
Sequence: 1 MRKEVTPMLNKNYKPGPQF.....EEQDKEMTSAKQHLFVRKN 192

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Issued_Patents_AA:*

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	DB ID	Description
1	7	3.6	132 4	US-09-134-001C-4894
2	7	3.6	220 4	US-09-297-937C-7
3	7	3.6	284 3	US-09-053-197A-2
4	7	3.6	284 4	US-09-085-761A-2
5	7	3.6	317 4	US-08-937-067-6
6	7	3.6	1234 2	US-08-317-310A-15
7	7	3.6	1234 5	PCT-US95-13041-15
8	7	3.6	1243 2	US-08-557-139-2
9	7	3.6	1276 4	US-09-297-937C-13
10	7	3.6	1321 2	US-08-317-310A-16
11	7	3.6	1321 2	US-08-317-310A-64
12	7	3.6	1321 5	PCT-US95-13041-16
13	6	3.1	24 6	5240706-21
14	6	3.1	25 4	US-09-177-249-217
15	6	3.1	38 1	US-08-118-270-213
16	6	3.1	38 5	PCT-US93-08528-213
17	6	3.1	101 2	US-08-710-749-6
18	6	3.1	105 3	US-08-407-165-3
19	6	3.1	108 2	US-08-710-749-22
20	6	3.1	108 2	US-08-710-749-23
21	6	3.1	108 2	US-08-710-749-24
22	6	3.1	108 2	US-08-710-749-25
23	6	3.1	108 2	US-08-710-749-26
24	6	3.1	119 1	US-08-680-726A-62
25	6	3.1	119 4	US-09-092-409-62
26	6	3.1	139 1	US-08-680-726A-66
27	6	3.1	139 4	US-09-092-409-66
Sequence 4894, Ap				
Sequence 7, Appl				
Sequence 2, Appl				
Sequence 2, Appl				
Sequence 6, Appl				
Sequence 15, Appl				
Sequence 15, Appl				
Sequence 2, Appl				
Sequence 13, Appl				
Sequence 16, Appl				
Sequence 64, Appl				
Sequence 16, Appl				
Patent No. 5240706				
Sequence 217, App				
Sequence 213, App				
Sequence 213, App				
Sequence 3, Appl				
Sequence 22, Appl				
Sequence 24, Appl				
Sequence 25, Appl				
Sequence 26, Appl				
Sequence 62, Appl				
Sequence 66, Appl				
Sequence 66, Appl				

28	6	3.1	149	1	US-08-602-010A-18	Sequence 18, Appl
29	6	3.1	149	1	US-08-680-726A-18	Sequence 18, Appl
30	6	3.1	149	4	US-09-092-409-18	Sequence 18, Appl
31	6	3.1	196	4	US-08-975-762-32	Sequence 32, Appl
32	6	3.1	196	4	US-08-975-762-32	Sequence 32, Appl
33	6	3.1	196	4	US-08-821-324-32	Sequence 32, Appl
34	6	3.1	196	4	US-09-295-028-32	Sequence 32, Appl
35	6	3.1	196	4	US-09-295-028-32	Sequence 32, Appl
36	6	3.1	196	4	US-09-106-582-32	Sequence 32, Appl
37	6	3.1	196	4	US-09-106-582-32	Sequence 32, Appl
38	6	3.1	198	4	US-09-134-001C-3459	Sequence 3459, Ap
39	6	3.1	244	4	US-09-134-001C-4784	Sequence 4784, Ap
40	6	3.1	271	1	US-08-252-9950-11	Sequence 11, Appl
41	6	3.1	271	2	US-08-834-108-11	Sequence 11, Appl
42	6	3.1	294	4	US-09-424-349A-3	Sequence 3, Appl
43	6	3.1	331	4	US-09-457-046B-59	Sequence 59, Appl
44	6	3.1	369	2	US-08-663-566A-4	Sequence 4, Appl
45	6	3.1	369	2	US-08-023-610-4	Sequence 4, Appl
46	6	3.1	369	2	US-08-288-065A-4	Sequence 4, Appl
47	6	3.1	369	2	US-08-362-240A-4	Sequence 4, Appl
48	6	3.1	369	4	US-08-804-372A-2	Sequence 2, Appl
49	6	3.1	369	5	PCT-US95-10245-4	Sequence 4, Appl
50	6	3.1	385	2	US-08-892-715-2	Sequence 2, Appl
51	6	3.1	385	2	US-09-145-947-2	Sequence 2, Appl
52	6	3.1	385	4	US-09-265-642-2	Sequence 2, Appl
53	6	3.1	410	1	US-08-471-033-40	Sequence 40, Appl
54	6	3.1	410	1	US-08-471-033-43	Sequence 43, Appl
55	6	3.1	410	2	US-08-471-044-40	Sequence 40, Appl
56	6	3.1	410	2	US-08-471-044-43	Sequence 43, Appl
57	6	3.1	410	2	US-08-463-483A-40	Sequence 40, Appl
58	6	3.1	410	2	US-08-463-483A-43	Sequence 43, Appl
59	6	3.1	410	2	US-08-471-046A-40	Sequence 40, Appl
60	6	3.1	410	2	US-08-471-046A-43	Sequence 43, Appl
61	6	3.1	410	2	US-08-470-566B-40	Sequence 40, Appl
62	6	3.1	410	2	US-08-470-566B-43	Sequence 43, Appl
63	6	3.1	410	2	US-08-469-334-40	Sequence 40, Appl
64	6	3.1	410	2	US-08-469-334-43	Sequence 43, Appl
65	6	3.1	410	3	US-09-300-529-40	Sequence 40, Appl
66	6	3.1	410	3	US-09-300-529-43	Sequence 43, Appl
67	6	3.1	413	4	US-08-942-572-2	Sequence 2, Appl
68	6	3.1	446	4	US-08-960-780-52	Sequence 52, Appl
69	6	3.1	446	4	US-09-073-898-52	Sequence 52, Appl
70	6	3.1	449	1	US-08-471-033-46	Sequence 46, Appl
71	6	3.1	449	1	US-08-471-044-46	Sequence 46, Appl
72	6	3.1	449	2	US-08-463-483A-46	Sequence 46, Appl
73	6	3.1	449	2	US-08-471-046A-46	Sequence 46, Appl
74	6	3.1	449	2	US-08-470-566B-46	Sequence 46, Appl
75	6	3.1	449	2	US-08-469-334-46	Sequence 46, Appl
76	6	3.1	449	3	US-09-300-529-46	Sequence 46, Appl
77	6	3.1	462	1	US-08-471-033-2	Sequence 2, Appl
78	6	3.1	462	2	US-08-471-044-2	Sequence 2, Appl
79	6	3.1	462	2	US-08-463-483A-2	Sequence 2, Appl
80	6	3.1	462	2	US-08-471-046A-2	Sequence 2, Appl
81	6	3.1	462	2	US-08-470-566B-2	Sequence 2, Appl
82	6	3.1	462	2	US-08-469-334-2	Sequence 2, Appl
83	6	3.1	462	3	US-09-300-529-2	Sequence 2, Appl
84	6	3.1	472	4	US-09-088-425-1	Sequence 1, Appl
85	6	3.1	482	4	US-09-134-001C-4309	Sequence 4309, Ap
86	6	3.1	516	2	US-08-676-166A-7	Sequence 7, Appl
87	6	3.1	520	4	US-09-326-203A-2	Sequence 2, Appl
88	6	3.1	525	4	US-08-676-166A-3	Sequence 3, Appl
89	6	3.1	600	6	5240706-1	Patent No. 5240706
90	6	3.1	603	4	US-09-198-122-2	Sequence 2, Appl
91	6	3.1	607	4	US-09-204-208A-11	Sequence 11, Appl
92	6	3.1	621	1	US-08-295-814E-6	Sequence 6, Appl
93	6	3.1	621	1	US-09-343-361-6	Sequence 6, Appl
94	6	3.1	621	5	PCT-US93-01959-6	Sequence 6, Appl
95	6	3.1	630	4	US-09-228-986-71	Sequence 71, Appl
96	6	3.1	650	4	US-09-232-191-29	Sequence 29, Appl
97	6	3.1	650	4	US-09-232-200-79	Sequence 79, Appl
98	6	3.1	650	4	US-09-232-200-97	Sequence 97, Appl
99	6	3.1	650	4	US-09-232-197-79	Sequence 79, Appl
100	6	3.1	650	4	US-09-232-197-97	Sequence 97, Appl

THIS PAGE BLANK (USPTO)


```

PR 04-OCT-2000; 2000GB-0024263.
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI; 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
PT
XX
XX Example 4; SEQ ID NO: 30040; 650pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is a protein encoded by one of
CC the probes of the invention.
XX
XX Sequence 377 AA;
SQ
XX
XX Query Match 9.1%; Score 91.5; DB 22; Length 377;
XX Best Local Similarity 28.1%; Pred. No. 0.64;
XX Matches 32; Conservative 14; Mismatches 41; Indels 27; Gaps 4;
XX
QY 78 YKD---ASTIRKNSRISRLLEDYIKEYCNFCAYFVLENNPNPRDIKFDDEPHKRRKRSRK 134
DB 147 YSDRSSESPRSRSRSRSYSRYTR-----SRSLASSHSRSR 186
XX
QY 135 SQSKSKSTRN---NRSQSNANAHFTSKRKDKTRRQERHIKEQDKEMTSKQKH 185
DB 187 SPSSSRHSRNKYSDHSQCSRSSSYTIS-SDDGRRAKRRLRSSGKKNVSHKKH 239
XX
XX RESULT 39
XX AAM70360
XX ID AAM70360 standard; Protein; 377 AA.
XX AC AAM70360;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human bone marrow expressed probe encoded protein SEQ ID NO: 30666.
XX
XX Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX
XX Homo sapiens.
XX
XX WO200157276-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00668.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI; 2001-488900/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
PT

```

```

XX
XX Example 4; SEQ ID NO: 30566; 658pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is a
CC protein encoded by one of the probes of the invention.
XX
XX Sequence 377 AA;
SQ
XX
XX Query Match 9.1%; Score 91.5; DB 22; Length 377;
XX Best Local Similarity 28.1%; Pred. No. 0.64;
XX Matches 32; Conservative 14; Mismatches 41; Indels 27; Gaps 4;
XX
QY 78 YKD---ASTIRKNSRISRLLEDYIKEYCNFCAYFVLENNPNPRDIKFDDEPHKRRKRSRK 134
DB 147 YSDRSSESPRSRSRSRSYSRYTR-----SRSLASSHSRSR 186
XX
QY 135 SQSKSKSTRN---NRSQSNANAHFTSKRKDKTRRQERHIKEQDKEMTSKQKH 185
DB 187 SPSSSRHSRNKYSDHSQCSRSSSYTIS-SDDGRRAKRRLRSSGKKNVSHKKH 239
XX
XX RESULT 40
XX AAM18201
XX ID AAM18201 standard; Protein; 377 AA.
XX AC AAM18201;
XX
XX 12-OCT-2001 (first entry)
XX
XX Peptide #4635 encoded by probe for measuring cervical gene expression.
XX
XX Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer.
XX
XX Homo sapiens.
XX
XX WO200157278-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00670.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI; 2001-488901/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
PT
XX
XX Claim 27; SEQ ID No 23027; 487pp; English.
XX
XX The present invention relates to human single exon nucleic acid probes
CC (SENPs: see AAI10068-AA128459). The present sequence is a peptide encoded
CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs
CC can be used to produce a single exon microarray, which can be used for
CC measuring human gene expression in a sample derived from human cervical
CC epithelial cells. By measuring gene expression, the probes are therefore
CC useful in grading and/or staging of diseases of the cervix, notably
CC cervical cancer.
XX

```


CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising
CC the nucleic acids encoding the polypeptides and cells genetically
CC engineered to express them are also useful for producing the proteins.
CC The proteins are useful in genetic vaccination, testing and
CC therapy, and can be used as nutritional supplements. They may be used to
CC increase stem cell proliferation; to regulate haematopoiesis; and in
CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
CC immune suppression and/or stimulation; as anti-inflammatory agents; and
CC in treatment of leukemias. AA029510-AA03304 represent the amino acid
CC sequences of novel human secreted proteins of the invention.

XX Sequence 583 AA;

Query Match 9.3%; Score 93.5; DB 22; Length 583;
Best Local Similarity 21.3%; Pred. No. 0.69;
Matches 51; Conservative 41; Mismatches 64; Indels 83; Gaps 13;

QY 17 GPQIFHFNIVKSDDTFQVIVNEKSAFDVTVGQRFSEILKY-----DFIVGDWGN 70

Db 92 GYGFVEDDLROADDAVYEL--NGKD-----LCGER---VIVEHARGPRD---GSYGSG 138

QY 71 Q-----LRLRGFYKDASTIRKNSRI-----SR-----LEDYK----- 98

Db 139 RSGYGYRRSGRQKYGPPTRTDRDLIVENLTSCSWQDLKDYMRQAGEVTVADAHKGRQKM 198

QY 99 -----EYCNFCGAYFVLE-----NPNRDIKFDDERP-----HKR----- 128

Db 199 KGVIEVSYSDMKRALEKLDGTEVNGKRLVLEDKPGSRRRRSYSRSHSRSRSHSRSHS 258

QY 129 RKSRSKSQSKSQTNRNRSOSNANAHTSK--KRDKTKRQERHKEEDKEMTSKQ 185

Db 259 RKSRSGSKSKSHSRSRSGSRSGSRSGSRSGSRSGSRSGSRSGSRSGSRSGSRSH 317

RESULT 31

ABG10283
ID ABG10283 standard; Protein; 737 AA.

XX AC ABG10283;

DT 13-FEB-2002 (first entry)

XX Novel human diagnostic protein #10274.

DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

XX WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR N-PSDB; AAS74470.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity

XX Claim 20; SEQ ID No 40642; 103pp; English.
PS The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 737 AA;

Query Match 9.2%; Score 93; DB 22; Length 737;
Best Local Similarity 26.1%; Pred. No. 1.1;
Matches 31; Conservative 25; Mismatches 27; Indels 36; Gaps 6;

QY 85 RKNRSIRSLDYKEYCNFCAYFVLEPNRDIKFDDEP-----HKRRKSRKSKS 137

Db 400 RKNREKKVKDH-----KSNSEKEDIRNSEKEDYKKNVKKRAKSKSRKS 446

QY 138 -SKSQTNRNRSOSNANAHTSKR---KDKTKRQERHKEEQ-----DKEMTSKQ 184

Db 447 KEKSKSKERDSKHNRN-----EEKRMFKELKGRDHENVKEKEKQSDSKGDKQERSRSE 501

RESULT 32

AAR13320
ID AAR13320 standard; Protein; 982 AA.

XX AC AAR13320;

DT 17-DEC-2001 (updated)

DT 22-OCT-1991 (first entry)

XX DE Murine Natural Killer receptor.

XX KW NK; cytotoxic drugs; tumour cell.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Protein 8..982

FT Peptide 1..7

XX /label= mature_NK_receptor

XX /label= partial_signal_peptide

PN USN7535206-N.

XX 09-JUL-1991.

XX PF 08-JUN-1990; 90US-0143578.

XX PR 08-JUN-1990; 90US-0535206.

XX (USSH) NAT INST OF HEALTH.

XX PI Ortaldo J, Young H, Anderson S;

DT 22-OCT-2001 (first entry)
XX Human polypeptide SEQ ID NO 3439.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
XX WO200153312-A1.
XX
XX 26-JUL-2001.
XX
XX 26-DEC-2000; 2000WO-US34263.
XX
XX 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
XX (HYTE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI: 2001-442253/47.
DR N-PSDB; AA159450.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Example 6; SEQ ID NO 3439; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AA38642-AA42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
XX Sequence: 1150 AA;
SQ

Query Match 9.5%; Score 95.5; DB 22; Length 1150;
Best Local Similarity: 18.6%; Pred. No. 1;
Matches 36; Conservative 39; Mismatches 64; Indels 55; Gaps 7;
7;

QY 32 IEFQVINEKSAFDVAVFGORFSEILLKYDFIVGDWNEQLRLRGYKDASTI----- 84
DB 753 LDYRLDENKHSFEVSLFAELFMELQOR-----DFG---VRI---YKSLSLPEKEDKK 800
QY 85 -----RKNRIRISRLDYIKFCYCNFGCAYEVLENPNPNRDIKFDDEPHK-----R 128
DB 801 KKKKSKDKDKKEDRDE-----TDEPKPKRKSGDDKDKEDRDKEDK 850

QY 129 RKRSKSSQSSKSTQTRNNRSQSNANAHFTSKK-----RKDTRRROERHIKESQDKE 178
DB 851 RKDDSKDDDEEDNNDQDEYDPMEEAEDEEDDRDEEMTKRDKRDINRYCKERPSSKD 910
QY 179 MTSAKOHLFLVRKN 192
DB 911 KEKEKTQMITINRD 924
RESULT 28
AAB40574
ID AAB40574 standard; Protein; 2518 AA.
XX
XX AAB40574;
AC
XX 08-FEB-2001 (first entry)
DT
XX Human ORFX ORF338 polypeptide sequence SEQ ID NO:676.
DE
XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antichyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX
XX Homo sapiens.
XX
XX WO200058473-A2.
XX
XX 05-OCT-2000.
XX
XX 31-MAR-2000; 2000WO-US08621.
XX
XX 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shinkets RA, Leach M;
PI
DR WPI: 2000-602362/57.
XX
XX N-PSDB; AAC74783.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
XX Claim 11; Page 784-790; 5507pp; English.
XX
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antichyroid; antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,

```
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
PI WPI: 2001-235357/24.
DR N-PSDB; AAH34397.
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers -
XX Claim 11: Page 7273-7275; 9803pp; English.
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate p
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX SQ Sequence 540 AA;
Query Match 9.5%; Score 95.5; DB 22; Length 540;
Best Local Similarity 18.6%; Pred. No. 0.39;
Matches 36; Conservative 39; Mismatches 64; Indels 55; Gaps 7;
QY 32 IEFOLVINEKSAFDVTYFGQRFSEILLKYDFIVGDWNEQLRLRGFYKDASTI----- 84
Db 143 LDYRLDNKEHSFEVSLFAELFNEMLQR-----DFG---VRI---YKSLLSLPEKEDK 190
QY 85 -----RKNRSIRLEDYIKEYCNFGCAYFVLEPNPRDIKFDDERPHK-----R 128
Db 191 EKDKKSKDERKKKEERDDE-----TDEPKPKRKSGDDKKEDRDERKEDK 240
QY 129 RKSRKSKSQSKQTRNNRSQSNANAHTSKK-----RKDKTKRQERHIKEEQDK 178
Db 241 RKDSKDDDETEEDNQDEYDPMEAEAEDEEDRDEEEMTKRDDKRDINRYCKERPSKD 300
QY 179 MTSAKQHLLFVRKN 192
Db 301 KEKEKTOMITINRD 314
RESULT 26
AAB92973
ID AAB92973 standard; Protein; 718 AA.
XX AAB92973;
XX 26-JUN-2001 (first entry)
XX Human protein sequence SEQ ID NO:11681.
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX Homo sapiens.
XX EP1074617-A2.
XX 07-FEB-2001.
XX 28-JUL-2000; 2000EP-0116126.
```

```
XX 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isojari T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyaama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI: 2001-318749/34.
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX Claim 8; SEQ ID 11681; 2537pp + CD ROM; English.
XX The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX SQ Sequence 718 AA;
Query Match 9.5%; Score 95.5; DB 22; Length 718;
Best Local Similarity 18.6%; Pred. No. 0.56;
Matches 36; Conservative 39; Mismatches 64; Indels 55; Gaps 7;
QY 32 IEFOLVINEKSAFDVTYFGQRFSEILLKYDFIVGDWNEQLRLRGFYKDASTI----- 84
Db 321 LDYRLDNKEHSFEVSLFAELFNEMLQR-----DFG---VRI---YKSLLSLPEKEDK 368
QY 85 -----RKNRSIRLEDYIKEYCNFGCAYFVLEPNPRDIKFDDERPHK-----R 128
Db 369 EKDKKSKDERKKKEERDDE-----TDEPKPKRKSGDDKKEDRDERKEDK 418
QY 129 RKSRKSKSQSKQTRNNRSQSNANAHTSKK-----RKDKTKRQERHIKEEQDK 178
Db 419 RKDSKDDDETEEDNQDEYDPMEAEAEDEEDRDEEEMTKRDDKRDINRYCKERPSKD 478
QY 179 MTSAKQHLLFVRKN 192
Db 479 KEKEKTOMITINRD 492
RESULT 27
AAM40294
ID AAM40294 standard; Protein; 1150 AA.
XX AAM40294;
XX
```


Db 639 NKRKKKKKEEKEEERRRKEEED 666

RESULT 22

AAB54149

ID AAB54149 standard; Protein; 306 AA.

AC AAB54149;

XX

DT 09-MAR-2001 (first entry)

XX

DE Human pancreatic cancer antigen protein sequence SEQ ID NO: 601.

XX

DE Human; pancreas; pancreatic cancer; pancreatic cancer antigen; detection; diagnosis; identification; cytostatic; neuroprotective; KW KW nontropic; immunomodulatory; relaxant; contraceptive; gynaecological; KW KW antiinflammatory; cardiant; gene therapy; chromosome mapping; KW KW linkage analysis; tissue identification; tissue typing; forensic; KW KW neural; immune system; muscular; reproductive; gastrointestinal; KW KW pulmonary; cardiovascular; renal; proliferative.

XX

OS Homo sapiens.

XX

PN WO200055320-A1.

XX

PD 21-SEP-2000.

XX

PF 08-MAR-2000; 2000WO-US05989.

XX

PR 12-MAR-1999; 99US-0124270.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Rosen CA, Ruben SM;

XX

DR WPI; 2000-579444/54.

DR N-PSDB; AAC98914.

XX

PT New nucleic acid that is a pancreatic cancer antigen for preventing, treating, or ameliorating a medical condition, particular pancreatic cancer, or for use in assays for diagnosing a pathological condition.

XX

PS Claim 11; Page 1037-1038; 1379pp; English.

XX

CC AAC98773 to AAC99231 encode the human pancreatic cancer associated proteins, called pancreatic cancer antigens, given in AAB54008 to AAB54466. The human pancreatic cancer antigens have cytostatic, neuroprotective, nontropic, immunomodulatory, relaxant, contraceptive, gynaecological, cardiant and antiinflammatory activities, and can be used in gene therapy. The polynucleotide and proteins can be used for preventing, treating, or ameliorating a medical condition or in assays for diagnosing a pathological condition or a susceptibility to one in a subject. Binding partners to the proteins and the activity of the proteins can be identified. The pancreatic cancer antigens can be used to detect, treat or prevent pancreatic disorders, especially cancer.

CC

CC Agonists and antagonists to the antigens can be screened for. The pancreatic cancer antigen polynucleotides can be used to design nucleic acid hybridisation probes that can be used in chromosome mapping, linkage analysis, tissue identification and/or typing and a variety of forensic and diagnostic methods. The proteins can be used to generate antibodies which are used to purify, detect and target the polypeptides, including both in vivo and in vitro diagnostic and therapeutic methods. The proteins can be used to treat or prevent neural, immune system, muscular, reproductive, gastrointestinal, pulmonary, cardiovascular, renal or proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent sequences used in the exemplification of the present invention.

XX

SQ Sequence 306 AA;

Query Match 9.58; Score 95.5; DB 21; Length 306;

Best Local Similarity 25.08; Pred. No. 0.19;

Matches 34; Conservative 22; Mismatches 45; Indels 35; Gaps 5;

Qy 78 YKDASTIRKNSRISRLDYIKCYNFCGAYFVLE-----NPNPRDIKFDDEP----- 125

Db 64 YADAHKGRKN-----EGVIEFVSYSMDKRALEKLDGTEVNGRKIRLVEDKPGSRRRS 116

Qy 126 -----HKR-----RKSRKSSKSSQTSRNNRSQSNANAHFTSK--KRKDTKRQER 169

Db 117 YSRSRSHSRSRSHSRSHSRSGSGSKSHSKSRSGSRSGSRSGSRSGSRSGSRSGSKK 176

Qy 170 HIKECDKEMTSKOH 185

Db 177 EKSRSFSKDKSRSRSH 192

RESULT 23

AAM41227

ID AAM41227 standard; Protein; 347 AA.

XX

AC AAM41227;

XX

DT 22-OCT-2001 (first entry)

XX

DE Human polypeptide SEQ ID NO 6158.

XX

DE Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer; KW KW peripheral nervous system; neuropathy; central nervous system; CNS; KW KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; KW KW amyotrophic lateral sclerosis; Shy-brager Syndrome; chemotactic; KW KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation; KW KW leukaemia.

XX

OS Homo sapiens.

XX

PN WO200153313-A1.

XX

PD 26-JUL-2001.

XX

PF 26-DEC-2000; 2000WO-US34263.

XX

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

PR 09-JUL-2000; 2000US-0598042.

PR 19-JUL-2000; 2000US-0620312.

PR 03-AUG-2000; 2000US-0653450.

PR 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.

PR 29-NOV-2000; 2000US-0727344.

XX

(HYSE-) HY SEQ INC.

XX

PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;

PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX

DR WPI; 2001-442253/47.

DR N-PSDB; AAI60383.

XX

PT Novel nucleic acids and polypeptides, useful for treating disorders such as central nervous system injuries -

XX

PS Example 2; SEQ ID NO 6158; 10078pp; English.

XX

CC The invention relates to human nucleic acids (AAI57798-AAI61369) and the encoded polypeptides (AAM38642-AAM42213) with nontropic, immunosuppressant and cytostatic activity. The polynucleotides are useful in gene therapy. A composition containing a polypeptide or polynucleotide of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nervous injuries, peripheral neuropathy and localised neuropathies and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the utilisation of the activities such as: immune system suppression, CC Activin/irhbin activity, chemotactic/chemokinetic activity, haemostatic CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,

PR	03-NOV-2000; 2000US-0706197.
XX	
PA	(ISIS-) ISIS PHARM INC.
PA	(COLD-) COLD SPRING HARBOR LAB.
XX	
PI	Bennett CF, Spector DL, Wyatt JR;
XX	
XX	WPI; 2002-479763/51.
DR	N-PSDB; AAD39557.
XX	
XX	Novel antisense compounds targeted to nucleic acids encoding SR-cyp.
PT	Clk-associated RS cyclophilin for modulating the gene expression and
PT	treating hyperproliferative disorders such as cancer
XX	
XX	Example 13; Page 95-99; 117pp; English.
XX	
XX	The invention relates to antisense compounds targetted to a nucleic
CC	acid molecule encoding human SR-cyp (Clk-associated RS cyclophilin)
CC	to inhibit its expression. SR-cyp is also referred to as CARS-cyp.
CC	Antisense compounds of the invention are used for treating diseases
CC	or conditions associated with SR-cyp. The diseases treated include
CC	hyperproliferative disorders e.g. cancer or hyperproliferative
CC	disorders resulting from an alternative splicing event. They are
CC	useful for diagnostics, therapeutics and as research reagents, e.g.
CC	prophylactically to prevent or delay infection, inflammation or
CC	tumour formation. They are also used in antisense therapy. The present
CC	sequence is human SR-cyp protein.
XX	
XX	Sequence 754 AA:
XX	

Query Match	9.8%;	Score 99;	DB 23;	Length 754;		
Best Local Similarity	21.7%;	Pred. No. 0.26;				
Matches 41;	Conservative 25;	Mismatches 51;	Indels 72;	Gaps		
QY	61	DFIV--GDW-----	NEOLRLRGFYKDA	TIKKNRSIRLEDY	KEYCN-----	FGCAY 107
		:	:	:	:	:
Db	71	DFWQGGDFSEGNCR	GESI-YCGF	FEDESFAV	HKHK-----	EFLLSMANRCKDINGSOF 124
QY	108	FVLEPNPR-----				-----DJK 119
		:	:			:
Db	125	FITTKPTPLDGHVV	FGVOISGQEVV	RVREIENQKTD	AAASKPFAEVRIL	SCGELIPKSKVK 184
QY	120	FDDERPHKRKR	SKSKS----	QSSKSQTNRNRS	OSNANAHEFTSK	RKDKTRQERHIKEE 174
		: :	: :	: :	: :	: :
Db	185	KEBKRRHKSSSS	SSSSSSSDSD	SSSSSQSSSDSD	SSSESATEEKS	KRKKKKHRRKNSRKHKE 244
QY	175	QDKEMTSAK 183				
		:	:			:
Db	245	KKKRKKSKK 253				

RESULT 17	
AAB41897	Human ORFX ORF1661 polypeptide sequence SEQ ID NO:3322.
ID	AAB41897 standard; Protein; 454 AA.
XX	
XX	
XX	AAB41897;
XX	
DT	08-FEB-2001 (first entry)
XX	
XX	
DE	
XX	
KW	Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW	vulnary; antiapoptotic; antiparkinsonian; nontropic; neuroprotective;
KW	anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW	immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW	hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW	antiviral; antibacterial; antifungal; antineumatic; antithyroid;
KW	antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW	neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW	cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW	cholesterol ester storage; systemic lupus erythematosus; infection;
KW	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW	

KW	bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW	thrombosis; contraceptive.
XX	
OS	Homo sapiens.
XX	
PN	WO200058473-A2.
XX	
PD	05-OCT-2000.
XX	
PF	31-MAR-2000; 2000WO-US08621.
XX	
PR	31-MAR-1999; 99US-0127607.
PR	02-APR-1999; 99US-0127636.
PR	05-APR-1999; 99US-0127728.
PR	30-MAR-2000; 2000US-0540763.
XX	
PA	(CURA-) CURAGEN CORP.
XX	
PI	Shimkets RA, Leach M;
XX	
DR	WPI; 2000-602362/57.
XX	N-PSDB; AAC76106.
XX	
PT	Novel nucleic acids and peptides derived from open reading frame X,
PT	useful for treating e.g. cancers, proliferative disorders,
PT	neurodegenerative disorders and cardiovascular disease -
XX	
PS	Chalm 11; Page 2511-2512; 5507pp; English.
XX	
CC	AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC	which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC	sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC	antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC	osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC	immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC	antidiabetic; hypotensive; dermatological; immunosuppressive;
CC	antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC	antithyroid; and antinaemic. The sequences can be used for determining
CC	the presence of or predisposition to, or preventing or treating
CC	pathological conditions associated with an ORFX-associated disorder. The
CC	nucleic acids can be used to express ORFX proteins in gene therapy
CC	vectors. The proteins and nucleic acids may be used to treat cancers,
CC	proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC	graft vs host disease, cardiovascular disease, diabetes mellitus,
CC	hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC	erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC	bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC	allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC	nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC	coagulation; to inhibit thrombosis; and as a contraceptive.
XX	
SO	Sequence 454 AA:

Query Match	9.7%;	Score 98;	DB 21;	Length 454;
Best Local Similarity	24.1%;	Pred. No. 0.17;		
Matches	47;	Conservative 31;	Mismatches 81;	Indels 36; Gaps 17;
Qy	7	PEMLNKNYPCQPQFHFNIVKSDIEFQLVINEKSAFDVTVFGRFSSEILLKYDVI	Gd 66	
Dg	172	PNATLGLPQPLMGVNDPSKIDEIRTVIVGNLSQTTT-----ADLLEFFKQVGE	225	
Qy	67	WGNEOLRLRGYPYKDASTRKNSRI-----SRLEDYIK---EYCYNCGAYFWLE	111	
Dg	226	VK-----FADGRLNHNSNAIVKPPTMPQAAAKEELVEVMKRVRAEQSFISAAI	276	
Qy	112	NPNPRDIKFEDDERPHKRKRSKSKOSSKSOTNNKFSQS--NANAHFTSKKR-KDFKRQE	168	
Dg	277	SGKSNERKGGRSHTRSKRSK----SSKSHSRKRSKHKHSRSHNRSSRKQDRRRS	333	
Qy	169	RHIKEEQDKEMTSAK	183	
Dg	334	PHKKRSKSRERRKSR	348	


```
Db 354 -VOIIMDIKKNAKKKKKKKKKKKKKKKK-----EKKKKKKKE 403
QY 169 RHIEQDQKEMTSKQ 184
Db 404 KNKKKEKKKKKKKE 419

RESULT 13
AAE15257
XX AAE15257 standard; Protein; 624 AA.
AC AAE15257;
XX
DT 07-MAR-2002 (first entry)
XX
DE Human RNA metabolism protein-20 (RMEP-20).
XX
KW Human; RNA metabolism protein-20; RMEP-20; gout; nervous system disorder;
KW autoimmune; inflammatory; cell proliferative; developmental; thyroiditis;
KW gene therapy; epilepsy; dementia; stroke; Alzheimer's disease; amnesia;
KW Parkinson's disease; prion disease; insomnia; endocrine disorder; AIDS;
KW Acquired Immune Deficiency Syndrome; mental disorder; allergy; anaemia;
KW asthma; atherosclerosis; Crohn's disease; rheumatoid arthritis; vaccine;
KW glomerulonephritis; multiple sclerosis; diabetes mellitus; osteoporosis;
KW cancer; cirrhosis; hepatitis; psoriasis; transgenic animal; antiulcer;
KW tranquiliser; drug screening; pancreatitis; renal tubular acidosis;
KW systemic lupus erythematosus; colitis.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 21..93
FT /label= RRM
FT /note= "RNA recognition motif"
FT Domain 184..253
FT /label= RRM
FT /note= "RNA recognition motif"
XX
WO200183524-A2.
XX
08-NOV-2001.
XX
27-APR-2001; 2001WO-US13862.
XX
28-APR-2000; 2000US-200184P.
PR 04-MAY-2000; 2000US-201875P.
PR 04-MAY-2000; 2000US-202050P.
PR 06-JUN-2000; 2000US-210232P.
PR 25-JUL-2000; 2000US-220553P.
XX
(INCY-) INCYTE GENOMICS INC.
XX
Lal P, Yue H, Tang YT, Lu DAM, Azimzai Y, Au-Young J, Hillman JL;
PI Baughn MR, Yao MG, Burford N, Batra S, Policky JJ;
XX
WPI; 2002-034502/04.
DR N-PSDB; AAD24393.
XX
PT New human RNA metabolism protein for diagnosing or treating nervous
PT system disorders, autoimmune/inflammatory disorders, cell proliferative
PT disorders and developmental disorders -
XX
Claim 1; Page 146-147; 196pp; English.
XX
The invention relates to human RNA metabolism proteins (RMEP) and their
CC corresponding cDNA molecules. RMEP and its DNA are used for diagnosing,
CC treating and preventing nervous system disorders (epilepsy, dementia,
CC stroke, Alzheimer's disease, Huntington's disease, Parkinson's disease);
CC prion diseases; fatal familial insomnia, nutritional and metabolic
CC diseases of the nervous system; inherited, metabolic, endocrine and
CC toxic myopathy; mental disorders (mood, anxiety, schizophrenic disorders)
CC amnesia and Tourette's disorder; autoimmune/inflammatory disorders (AIDS-
```

```
CC acquired immune deficiency syndrome, allergies, anaemia, asthma, gout,
CC atherosclerosis, Crohn's disease, diabetes mellitus, glomerulonephritis,
CC Hashimoto's thyroiditis, multiple sclerosis, osteoarthritis, rheumatoid
CC arthritis, osteoporosis, pancreatitis, systemic lupus erythematosus,
CC ulcerative colitis, and infections); cell proliferative disorders (cancer
CC arteriosclerosis, cirrhosis, hepatitis, psoriasis); and developmental
CC disorders (renal tubular acidosis). RMEP DNA is useful in drug screening
CC techniques, gene therapy and for creating transgenic animals. The present
XX sequence is human RMEP-20 protein.
SQ Sequence 624 AA;
Query Match 10.0%; Score 101; DB 23; Length 624;
Best Local Similarity 27.4%; Pred. No. 0.13;
Matches 43; Conservative 26; Mismatches 68; Indels 20; Gaps 6;
QY 32 IEF--QLVINEKSAFDVTFVGQRFSEILLKYDFIVGQNEQLRGFYKDASTIRKNSR 89
Db 227 VEFADONSVPRALAFNGVFGDRPLKINHNNNAIVRP--PEMTPOAAAKELEEVKRR-- 281
QY 90 ISRLDYIKYCNFGCAYFVLENPNDIKFDDERPHKRKSKSKSSQSSKTSQTRNNRSQS 149
Db 282 -----VREAQSFISAAIEPESGKSNERKGRSGRSHTRSKRSKRS---SSKSHSRKRSQS 331
QY 150 --NANAHFTSKKR-KDTRKQERHKEEQDKEMTSK 183
Db 332 KHRSRHNRSSRSQKORRRKSKSPHKRKRKRERRKSR 368

RESULT 14
AAB41766
ID AAB41766 standard; Protein; 334 AA.
XX
AC AAB41766;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF1530 polypeptide sequence SEQ ID NO:3060.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX
OS Homo sapiens.
XX
WO200058473-A2.
XX
05-OCT-2000.
XX
31-MAR-2000; 2000WO-US08621.
XX
31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
(CURA-) CURAGEN CORP.
XX
Shimkets RA, Leach M;
PI
XX
WPI; 2000-602362/57.
DR N-PSDB; AAC75975.
```

```
XX SQ Sequence 604 AA;
Query Match 10.4%; Score 105; DB 22; Length 604;
Best Local Similarity 22.1%; Pred. No. 0.048;
Matches 36; Conservative 34; Mismatches 65; Indels 28; Gaps 5;

QY 46 VVFGQRTSEILLKYD---FTVGDWGNOLRLRGFYKDASTIRKNS-----RISRLDY 96
DB 218 LTOLKESFOOVLEDERSEHLKGERARWQQRMKMSQGEICTLKKEKODMRVVELEERS 277
QY 97 IKEYCNFGCAYEVLEN-----PNPRDIKFDDERPHKRRKRSKSSKQSTRNNR--S 147
DB 278 LSK-----LKNQMAELPPPEPAVPSEVLEQLHLRKELEVERAGELQSQVKNQHLS 327
QY 148 QSNANAHFTSKRKDTKRQRERHKEEDQKMTSAKQHLFFVR 190
DB 328 LLNRQERIRQERLRKQERLQEQHEKRLQALAKPHCAFSR 370

RESULT 11
AAV22206
ID AAV22206 standard; Protein; 361 AA.
XX AC AAV22206;
XX DT 13-SEP-1999 (first entry)
XX DE Biorhythm marker protein.
XX KW Biorhythm marker gene; circadian rhythm.
XX OS Rattus sp.
XX PN JP11169185-A.
XX PD 29-JUN-1999.
XX PF 12-DEC-1997; 97JP-0362890.
XX PR 12-DEC-1997; 97JP-0362890.
XX PA (AGEN ) AGENCY OF IND SCI & TECHNOLOGY.
XX PS (SAKA ) OTSUKA PHARM CO LTD.
XX DR WPI; 1999-422623/36.
XX PT N-PSDB; AAX84569.
XX PT Structure and function of a new biorhythm marker gene - and protein
XX PT encoded by it
XX PS Claim 7; Page 14-16; 23pp; Japanese.
XX CC This sequence is the biorhythm marker protein of the invention. The
XX CC protein is also referred to as a circadian rhythm protein.
XX SQ Sequence 361 AA;
Query Match 10.2%; Score 102.5; DB 20; Length 361;
Best Local Similarity 33.0%; Pred. No. 0.045;
Matches 30; Conservative 15; Mismatches 33; Indels 13; Gaps 4;

QY 103 FGCA-----YFVLENPNPRDIKFDDERPHKRRKRSKSSKQSTRNNRSQ--NANA 153
DB 35 FNCSELTQSGFFLPESGSKNERKGRSHTSRKRS---SSKSHSRKRSKRSKRSRS 91
QY 154 HFTSKKR-KDTKRRQERHKEEQDKMTSAK 183
DB 92 HNRSRQKDRRSKSPHKRSKSRERKSR 122

RESULT 12
AAB18187
ID AAB18187 standard; Protein; 508 AA.
XX AC AAB18187;
XX DT 07-NOV-2000 (first entry)
XX DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:44.
XX KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
XX KW antimalarial; malaria; protozoacide; infection; insecticide.
XX OS Plasmodium falciparum.
XX PN WO200025728-A2.
XX PD 11-MAY-2000.
XX PF 05-NOV-1999; 99WO-US26796.
XX PR 05-NOV-1998; 98US-0107131.
XX PA (HOFF/) HOFFMAN S.
XX PA (CARU/) CARUCCI D.
XX PA (GARD/) GARDNER M.
XX PA (VENT/) VENTER J C.
XX PI Hoffman S, Carucci D, Gardner M, Venter JC;
XX DR WPI; 2000-355347/31.
XX PT Proteins encoded by chromosome 2 of the human malarial parasite,
XX PT Plasmodium falciparum, usefull as antimalarial vaccines and in the
XX PT diagnosis of P.falciparum infection -
XX PS Disclosure: Page 108-110; 577pp; English.
XX CC The present invention describes proteins and their fragments (I) encoded
XX CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
XX CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
XX CC vaccines against P. falciparum infection comprising (I) or (II).
XX CC (1) and (II) are usefull for the development of vaccines against
XX CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
XX CC antibody raised to immunogens comprising the sequences of (I), are
XX CC usefull in the detection of infection with P. falciparum. Furthermore,
XX CC (I) (especially when they are rifins or secreted or membrane proteins)
XX CC can aid the identification of drugs to treat or prevent P. falciparum
XX CC infection, or they can be used to identify drug resistance in
XX CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
XX CC subsequent identification of proteins encoded by it will help to expand
XX CC our understanding of parasite biology, a process hampered by the
XX CC complexity of the parasitic lifecycle, and provide new targets for
XX CC vaccine and drug development. Parasite resistance to drugs and mosquito
XX CC resistance to insecticides have led to a resurgence of malaria in many
XX CC parts of the world, and there is a pressing need for vaccines and new
XX CC drugs. AAV70078 to AAV70287 and AAB18144 to AAB18352 represent nucleotide
XX CC and protein sequences given in the present invention, but which are not
XX CC specifically mentioned within the specification.
XX SQ Sequence 508 AA;
Query Match 10.0%; Score 101; DB 21; Length 508;
Best Local Similarity 20.9%; Pred. No. 0.099;
Matches 41; Conservative 38; Mismatches 61; Indels 56; Gaps 9;

QY 5 VTPENLNVNKYPGQFIHFENIV-----KSDDIE----FOLVINEKSAFDTVFG 50
DB 264 ITKE-CVWKYD-----DNIDHKQHKREKKSNIENMISHIYEKEQSHDIC--- 312
QY 51 QRFSBILKYDFIVGDWGNQELRLRGFYKDASTIRKRSRLRDY--IKEYCNFGCAYF 108
DB 313 -----NVLEENKEEKYNNLQKDVITNCNNDKV-KLEEYHHEKELNN----- 353
QY 109 VLENPNPRDIKFDDERPHKRRKRSKSSKQSTRNNRSQSNANAHFTSKRKDKTRQEE 168
```


Query Match	25.2%; Score 254.5; DB 23; Length 132;
Best Local Similarity	45.6%; Pred. No. 3.3e-18;
Matches	47; Conservative 24; Mismatches 29; Indels 3; Gaps
QY	24 ENIVKSVDIEFQIVINEKSAFDVTVFGQRFSEILLKYDFITVGWGNQQLRLRCGYKDAST 83
DB	2 ENMIKVQQQVFELIEEYRECDFEISFARSVDILDKFYVVGDIYDQLRLKLGFYKDSNK 61
QY	84 IRK-NSISRLSDYIKEYCNFGCAYFVLENPDRDI--KFDDE 123
DB	62 KAEISKYFSISQIDYILLECYCNFGCGCFVVRRLSPNEFIEEIDDK 104
RESULT 8	
AAG82114	
ID	AAG82114 standard; Protein; 129 AA.
XX	AAG82114;
XX	03-SEP-2001 (first entry)
DE	S. epidermidis open reading frame protein sequence SEQ ID NO:1322.
DE	Staphylococcus epidermidis SRI strain; infection; diagnosis;
KW	vaccination; endocarditis.
XX	Staphylococcus epidermidis.
OS	WO200134809-A2.
PN	17-MAY-2001.
XX	09-NOV-2000; 2000WO-US30782.
XX	09-NOV-1999; 99US-0164258.
PR	(GLAX) GLAXO GROUP LTD.
XX	Kimmerly WJ;
XX	WPI; 2001-316495/33.
DR	N-PSDB; AAH52964.
XX	Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT	useful for vaccinating against infections, e.g. endocarditis -
XX	Claim 18: Fage 376-377; 2189pp; English.
PS	AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
XX	(II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
CC	(I) and (II) can have antibacterial activity and therefore can be used
CC	in vaccination. The nucleic acids (I) may be used to produce the
CC	S. epidermidis polypeptides (II) via the production of vectors
CC	containing them which are used to produce hosts cells which express the
CC	polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC	used to vaccinate subjects and to raise antibodies against the bacteria.
CC	The polypeptides may also be used to assay for other inhibitors of their
CC	activity and therefore identify compounds that may be used for the
CC	treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC	AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC	polynucleotide sequences from the present invention. AAH55091 to
CC	AAH55098 represent oligonucleotide sequences and primers which are used
CC	in the exemplification of the present invention.
CC	N.B. The present invention specifically claims all the polynucleotide
CC	sequences given in the sequence listing of the present specification,
CC	however the sequence listing only goes up to SEQ ID NO:4454 so even
CC	though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC	no sequences are present for SEQ ID NO:4455 to 4464.
XX	Sequence 129 AA;
XX	Sequence 129 AA;

XX 02-JUL-2002 (first entry)
XX Streptococcus polypeptide SEQ ID NO 150.
XX
XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial;
KW antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX
XX Streptococcus pyogenes.
XX
XX WO200234771-A2.
XX
XX 02-MAY-2002.
XX
XX 29-OCT-2001; 2001WO-GB04789.
XX
XX 27-OCT-2000; 2000GB-0026333.
PR 24-NOV-2000; 2000GB-0028727.
PR 07-MAR-2001; 2001GB-0005640.
XX
XX (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
XX Telford J, Massignani V, Margarit Ros YI, Grandi G, Fraser C;
PI Tettelin H;
XX
XX WPI: 2002-352536/38.
DR N-PSDB: ABN66118.
XX
XX New Streptococcus protein for the treatment or prevention of infection
PT or disease caused by Streptococcus bacteria, such as meningitis, and
PT for detecting a compound that binds to the protein -
XX
XX Claim 1; Page 3168; 4525pp; English.
XX
XX The invention relates to a protein (ABP25413-ABP30895) from group B
CC Streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins.
XX
XX Sequence 214 AA;
XX
XX Query Match 55.4%; Score 558.5; DB 23; Length 214;
XX Best Local Similarity 51.7%; Pred. No. 4.3e-49;
XX Matches 107; Conservative 42; Mismatches 41; Indels 17; Gaps 4;
OY 1 MRKEVTPMLNKNYPGQFTHFENIVKSDDDIEFQVINEKSAFDVTVFGORFSEILLKY 60
DB 9 MKKELSPMYNKNYPGPKFTHFEQVRAEGIDLLLEDVKNADFDTSGQRYTEVLKY 68
OY 61 DFIVGWCNEQLRLRGFYKDSASTIRKNSRISLEDEYIKFCNFCAYFVLENPNRDIKF 120
DB 69 DYIVGWCNEQLRLRGFYKDSIDDKTKNRISLEDEYIKFCNFCAYFVLENLHPQDIKF 128
OY 121 DDERPHKRKRS-RSKS-----SSKSQTNNRSQSNANAHFTSKRKKDKTR 165
DB 129 EEPORPRKRSKPSKSNRRKNPNYSNOQPATPKSKS-RASKEQOPENQAFTSQRRSNTK 187
OY 166 RQERHIKEQDKEMTSAKOHLLFVRKN 192

DB 188 HKEKS-KRNOTSQLNTKISHFIIRKDD 213
XX
XX RESULT 4
XX AAU37821
ID AAU37821 standard; Protein; 176 AA.
XX
XX AC AAU37821;
XX
XX DT 14-FEB-2002 (first entry)
XX
XX DE Streptococcus pneumoniae cellular proliferation protein #250.
XX
XX KW Antisense: prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
XX OS Streptococcus pneumoniae.
XX
XX PN WO200170955-A2.
XX
XX PD 27-SEP-2001.
XX
XX PF 21-MAR-2001; 2001WO-US09180.
XX
XX PR 21-MAR-2000; 2000US-191078P.
XX 23-MAY-2000; 2000US-206848P.
XX 26-MAY-2000; 2000US-207727P.
XX 23-OCT-2000; 2000US-242578P.
XX 27-NOV-2000; 2000US-253625P.
XX 22-DEC-2000; 2000US-257931P.
XX 16-FEB-2001; 2001US-269308P.
XX
XX (ELIT-) ELJTRA PHARM INC.
XX
XX PA Haselbeck H, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX PI Yamamoto RJ, Xu HH;
XX
XX DR WPI: 2001-611495/70.
XX N-PSDB: AA55680.
XX
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Example 3; Seq ID No 13414; 511pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 176 AA;
XX
XX Query Match 44.8%; Score 452.5; DB 22; Length 176;
XX Best Local Similarity 47.6%; Pred. No. 2.4e-38;
XX Matches 91; Conservative 35; Mismatches 50; Indels 15; Gaps 4;
OY 1 MRKEVTPMLNKNYPGQFTHFENIVKSDDDIEFQVINEKSAFDVTVFGORFSEILLKY 60

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2003, 09:44:37 ; Search time 84 Seconds
(without alignments)
304.573 Million cell updates/sec

Title: US-09-868-352-23

Perfect score: 1009

Sequence: 1 MRKEVTPPEMLNKNYPGPQF.....EQDKEMTSAKQHLLFVRKN 192

Scoring table: BLOSUM62

Gapop 10.0 , Capext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1009	100.0	192	21	AA196805
2	980	97.1	192	23	AB25486
3	558.5	55.4	214	23	AB25487
4	452.5	44.8	176	22	AAU37821
5	452.5	44.8	176	22	AAU37974
6	331.5	32.9	251	23	AB53482
7	254.5	25.2	132	23	AB40049
8	243.5	24.1	129	22	AA282114
9	239.5	23.7	61	22	AB50067
10	105	10.4	604	23	ABG15166

11	102.5	10.2	361	20	AAV22206	Biorhythm marker p
12	101	10.0	508	21	AAE18187	Plasmodium falci
13	101	10.0	624	23	AAE15257	Human RNA metaboli
14	99	9.8	334	21	AAE41766	Human ORFX ORF1530
15	99	9.8	376	21	AAE43954	Human cancer assoc
16	99	9.8	754	23	AAE24596	Human SR-cyp prote
17	98	9.7	454	21	AAE41897	Human ORFX ORF1661
18	97.5	9.7	336	22	ABG68960	Drosophila melanog
19	96	9.5	323	22	ABG39441	Human polypeptide
20	96	9.5	735	22	ABG26213	Human human diagno
21	96	9.5	1074	22	ABG00401	Novel human diagno
22	95.5	9.5	306	21	AAE54149	Human pancreatic c
23	95.5	9.5	347	22	AAE41227	Human polypeptide
24	95.5	9.5	540	22	ABE11354	Human Ser/Arg rich
25	95.5	9.5	540	22	AAE74992	Human colon cancer
26	95.5	9.5	718	22	ABG2973	Human protein sequ
27	95.5	9.5	1150	22	AAE40294	Human polypeptide
28	95	9.4	2518	21	AAE40574	Human ORFX ORF338
29	94	9.3	242	22	ABG07223	Novel human diagno
30	93.5	9.3	583	22	AAU30814	Novel human secret
31	93	9.2	737	22	ABG10283	Novel human diagno
32	93	9.2	982	12	AAE13320	Murine Natural Kil
33	92	9.1	215	22	ABG26528	Novel human diagno
34	91.5	9.1	323	22	ABG14082	Novel human diagno
35	91.5	9.1	377	22	ABG31983	Peptide #4634 enco
36	91.5	9.1	377	22	ABG37226	Peptide #4732 enco
37	91.5	9.1	377	22	ABG22527	Protein #4526 enco
38	91.5	9.1	377	22	AAE57935	Human brain expres
39	91.5	9.1	377	22	AAE70360	Human bone marrow
40	91.5	9.1	377	22	AAE18201	Peptide #4635 enco
41	91.5	9.1	377	22	AAE30696	Peptide #4733 enco
42	91.5	9.1	377	22	AAE05818	Peptide #4500 enco
43	91.5	9.1	377	23	ABG40005	Human peptide enco
44	91.5	9.1	776	22	ABG29124	Novel human diagno
45	91.5	9.1	962	22	ABG04842	Novel human diagno

ALIGNMENTS

RESULT 1	
AA196805	
ID	AA196805 standard; Protein: 192 AA.
XX	
AC	AA196805;
XX	
DT	26-SEP-2000 (first entry)
XX	
DE	Streptococcus agalactiae pho3-1 protein.
XX	
KW	Group B Streptococcus; pho3-1; virulence; vaccine; uropathic;
KW	gene therapy; veterinary; mastitis; immunostimulant; antibacterial.
XX	
OS	Streptococcus agalactiae.
XX	
PN	WO200037646-A2.
XX	
PD	29-JUN-2000
XX	
PF	22-DEC-1999 99WO-GB04377.
XX	
PR	22-DEC-1998 98GB-0028345.
PR	22-DEC-1998 98GB-0028349.
PR	22-DEC-1998 98GB-0028350.
PR	22-DEC-1998 98GB-0028352.
PR	22-DEC-1998 98GB-0028353.
PR	22-DEC-1998 98GB-0028354.
PR	22-DEC-1998 98GB-0028355.
PR	22-DEC-1998 98GB-0028356.
PR	22-DEC-1998 98GB-0028357.
PR	22-DEC-1998 98GB-0028359.
PR	04-JAN-1999 99GB-0000082.
PR	04-JAN-1999 99GB-0000083.

Db 2152 ---ENQRHQRDRSRASRSLGRYTDVDTGLGTLDSMTTQSGDLPSPKRDQGRPK 2208

QY 162 DTKRQERH 170

Db 2209 DRKHROHH 2217

RESULT 39

US-09-134-001C-3930

; Sequence 3930, Application US/09134001C

; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C

; CURRENT FILING DATE: 1998-08-13

; PRIOR FILING DATE: 1998-06-13

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674

; SEQ ID NO 3930

; LENGTH: 564

; TYPE: PRT

; ORGANISM: Staphylococcus epidermidis

US-09-134-001C-3930

Query Match 7.7%; Score 77.5; DB 4; Length 564;

Best Local Similarity 22.3%; Pred. No. 4.3;

Matches 44; Conservative 36; Mismatches 60; Indels 57; Gaps 9;

QY 11 NTNKPQGFQFIHFENIVKSDDIIEFQLVINEKSAFDVTVFGQRFSEILLKYDFIVGDMGNE 70

Db 152 NANRY---TFVWKKSI---QNHESRMNENSKALYHELVIKTIPEIKKDHD---NDLTKE 202

QY 71 QLRLRGFYKD-----ASTIRKNSRISRLIEDYIKCYNFCGAYFVLENPNR 116

Db 203 EIDLIGSHLDKEIDLNQHDNEKTKIRKQIRLKRTK--IKKYK----- 246

QY 117 DIKFDEPHKRRKRSKSSQSKSOTRNNRSQSNANAHFTSKRKDKTKRQERHIKEOD 176

Db 247 --QIND---YSQRKHKEVQKSIKDRNSYSKTDHDTAF-----HRMKEDHMKNGOL 293

QY 177 K-----EMTSKQHLIF 188

Db 294 KPGYNLQIATNSQFVLF 310

RESULT 40

US-09-356-952-5

; Sequence 5, Application US/09356952

; Patent No. 6117663

; GENERAL INFORMATION:

; APPLICANT: Boriack-Stodin, Ann

; APPLICANT: Margarit, S. M.

; APPLICANT: Bor-Sogli, Dafna

; APPLICANT: Cole, Philip

; APPLICANT: Kuriyan, John

; TITLE OF INVENTION: A CRYSTAL OF A RAS-SOS COMPLEX AND METHODS OF USE

; TITLE OF INVENTION: THEREOF

; FILE REFERENCE: 600-1-228N

; CURRENT APPLICATION NUMBER: US/09/356,952

; CURRENT FILING DATE: 1999-07-19

; EARLIER APPLICATION NUMBER: 60/093,631

; EARLIER FILING DATE: 1998-07-21

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: PatentIn ver. 2.0

; SEQ ID NO 5

; LENGTH: 1048

; TYPE: PRT

; ORGANISM: Saccharomyces cerevisiae

US-09-356-952-5

Query Match

Best Local Similarity 7.7%; Score 77.5; DB 3; Length 1048;

Matches 43; Conservative 25; Mismatches 66; Indels 47; Gaps 7;

QY 8 EMLNKNYKPGPOFIHFENIVKSDDIIEFQLVINEKSAFDVTVFGQRFSEILL---KYDFI 63

Db 872 QSLNKLMDPKKNFIYRNELKS-----LHSAPCVPFEGVYLSDLTFTTDSGNPDYL 921

QY 64 VGDWGNQLRLRGFYKDASTIRKNSRISRLIEDYIKCYCNF-----GCAYEV 109

Db 922 VLEHG-----LKGVDHKKYINFNKR-SRLVDILQELIYFKKTHYDFTKDRTVIECISNS 975

QY 110 LENPNPDKIFD-----DERPHKRRKRSKSSQSKSOTRNNRSQSNANAHFTSKRKDKTK 164

Db 976 LENIPHIKQYQLSLIIEFKPRKKVVPNSNS-----NNKSQEKSRDDQDDEGKTSTK 1027

QY 165 R 165

Db 1028 K 1028

Search completed: February 26, 2003, 09:55:08

Job time : 25 secs

```
; FILING DATE: 30-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/482,384
; FILING DATE: 20-FEB-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/603,751
; FILING DATE: 04-APR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US89/01408
; FILING DATE: 04-APR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/176,899
; FILING DATE: 04-APR-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L.
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6362-55038
; TELEPHONE: (619) 238-0999
; TELEFAX: (619) 238-0062
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2265 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; FEATURE:
; OTHER INFORMATION: /product= "Alpha1A-2 subunit of
; OTHER INFORMATION: human calcium channel"
; US-08-149-097D-36

Query Match 7.8%; Score 78.5; DB 2; Length 2265;
Best Local Similarity 25.6%; Pred. No. 22;
Matches 33; Conservative 15; Mismatches 38; Indels 43; Gaps 5;

QY 69 NEQLRLRGFKYKDASTIRKNS-----RISRLDYIKYCNFCGAYFVLENPNRD 117
   |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 2105 NORRRGRPRGNLSTISDTSPMKRSASVLGPKARRLDY-----SLERVPE- 2151
   ::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
QY 118 IKFDDERPHKRRRSKSKSQTR-----NNRSQSNANAHTSKKR-----K 161
   ::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 2152 --ENQRHHQRRDRSRASRSLGRYTDVDTGLTGLTDLMTTQSGDLPSKERDQGRPK 2208
   |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
QY 162 DTKRRQERH 170
   |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 2209 DRKHQHHH 2217
   |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|

RESULT 38
US-08-149-097D-35
; Sequence 35, Application US/08149097D
; Patent No. 5874236
; GENERAL INFORMATION:
; APPLICANT: Harpold, Michael
; APPLICANT: Ellis, Steven
; APPLICANT: Williams, Mark
; APPLICANT: Feldman, Daniel
; APPLICANT: McCue, Ann
; APPLICANT: Brenner, Robert
; TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND
; TITLE OF INVENTION: METHODS
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: California
; COUNTRY: USA
```

```
; ZIP: 92101-2926
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/149,097D
; FILING DATE: 05-NOV-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/105,536
; FILING DATE: 11-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US92/06903
; FILING DATE: 14-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/914,231
; FILING DATE: 13-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/868,354
; FILING DATE: 10-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/745,206
; FILING DATE: 15-AUG-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/620,250
; FILING DATE: 30-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/482,384
; FILING DATE: 20-FEB-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/603,751
; FILING DATE: 04-APR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US89/01408
; FILING DATE: 04-APR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/176,899
; FILING DATE: 04-APR-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L.
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6362-55038
; TELEPHONE: (619) 238-0999
; TELEFAX: (619) 238-0062
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2509 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; FEATURE:
; OTHER INFORMATION: /product= "Alpha1A-1 subunit of
; OTHER INFORMATION: human calcium channel"
; US-08-149-097D-35

Query Match 7.8%; Score 78.5; DB 2; Length 2509;
Best Local Similarity 25.6%; Pred. No. 25;
Matches 33; Conservative 15; Mismatches 38; Indels 43; Gaps 5;

QY 69 NEQLRLRGFKYKDASTIRKNS-----RISRLDYIKYCNFCGAYFVLENPNRD 117
   |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 2105 NORRRGRPRGNLSTISDTSPMKRSASVLGPKARRLDY-----SLERVPE- 2151
   ::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
QY 118 IKFDDERPHKRRRSKSKSQTR-----NNRSQSNANAHTSKKR-----K 161
   ::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
```



```
; APPLICANT: He, Tong-Chuan
; APPLICANT: Kinzler, Kenneth
; APPLICANT: Vogelstein, Bert
; TITLE OF INVENTION: Beta Catenin, TCF-4, and APC Interact to
; TITLE OF INVENTION: Prevent Cancer
; FILE REFERENCE: 1107.75741
; CURRENT APPLICATION NUMBER: US/09/136,605A
; CURRENT FILING DATE: 1998-08-20
; EARLIER APPLICATION NUMBER: 08/821,355
; EARLIER FILING DATE: 1997-03-20
; EARLIER APPLICATION NUMBER: 09/003,687
; EARLIER FILING DATE: 1998-01-06
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 2973
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-136-605-7

Query Match      7.9%; Score 79.5; DB 4; Length 2973;
Best Local Similarity 23.5%; Pred. No. 25;
Matches 28; Conservative 16; Mismatches 42; Indels 33; Gaps 3;

QY 78 YKDASTIRKNSRISRLDYIKYCNFCGAYFVLENPNPRDIKEDDERPH-----126
Db 1135 YEDDKPTNTYSERYSEIEQHEE-----ERPTNYSIKYNEERKHVDQPIDYSLKY 1183
QY 127 -----KRRKSRKSSQSKSQTNRNRSQSNANAHFTSKRKDTRRQERHIKEQDK 177
Db 1184 ATDIPSSQKQSFSSKSSQSKSTHEHSSSEN---TSTPSSNAKRNQLHPSSAQSR 1239

RESULT 33
US-08-698-407-4
; Sequence 4, Application US/08698407
; Patent No. 5856128
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Au-Young, Janice
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: A NOVEL HUMAN NUCLEIC ACID BINDING PROTEIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/698,407
; FILING DATE: Filed Herewith
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0116 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 475 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 267188
; US-09-136-605-7

Query Match      7.9%; Score 79.5; DB 4; Length 2973;
Best Local Similarity 23.5%; Pred. No. 25;
Matches 28; Conservative 16; Mismatches 42; Indels 33; Gaps 3;
```

```
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 267188
; US-08-698-407-4

Query Match      7.8%; Score 79; DB 2; Length 475;
Best Local Similarity 33.8%; Pred. No. 2.3;
Matches 23; Conservative 12; Mismatches 21; Indels 12; Gaps 3;

QY 111 ENPNPRDIK-FDDERPHKRRKRSKSSQSKSQTNRNRSQSNANAHFTSKRKDKTKRQE- 168
Db 13 ENKQERDKENRHRKRSRSDRDKRRSRDRNRNDQRSAS-----RDRRRRSKP 64
QY 169 --RHIKEE 174
Db 65 LTRGAKEE 72

RESULT 34
US-09-195-855-4
; Sequence 4, Application US/09195855
; Patent No. 6015788
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Au-Young, Janice
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: A NOVEL HUMAN NUCLEIC ACID BINDING PROTEIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/195,855
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/698,407
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0116 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 475 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 267188
; US-09-195-855-4

Query Match      7.8%; Score 79; DB 3; Length 475;
Best Local Similarity 33.8%; Pred. No. 2.3;
Matches 23; Conservative 12; Mismatches 21; Indels 12; Gaps 3;

QY 111 ENPNPRDIK-FDDERPHKRRKRSKSSQSKSQTNRNRSQSNANAHFTSKRKDKTKRQE- 168
Db 13 ENKQERDKENRHRKRSRSDRDKRRSRDRNRNDQRSAS-----RDRRRRSKP 64
```



```

;
; TELEPHONE 202 508 9100
; TELEFAX: 202 508 9299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3843 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-370-235A-2
;
; Query Match 7.9%; Score 79.5; DB 2; Length 2843;
; Best Local Similarity 23.5%; Pred. No. 23;
; Matches 28; Conservative 16; Mismatches 42; Indels 33; Gaps
;
QY 78 YKDASTIRKNSIRISLEDIYIKCYNFCGAYFVLENPNDRIKFDDEPH----- 126
;
DB 1135 YEDDRPNFVNSERYSEEQHEE-----EPTNYSKYNEEKRHVDQIDYSLKY 1183
;
QY 127 -----KRRKRSKSSQSSKSQTNRNRSQSNANAHFTSKKRKDKTKRKRQERHIKEQDK 177
;
DB 1184 ATDIPSSQKQSFSEKSSGQSSKTEHMSSSSEN--TSTPSSNAKRNQLHPSSAQSR 1239
;
;
RESULT 27
; US-08-450-582-2
; Sequence 2, Application US/08450582
; Patent No. 6114124
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERTIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bannet & Witcoff, Ltd.
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20301-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/450,582
; FILING DATE:
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/452,655
; FILING DATE: 25-MAY-1995
; APPLICATION NUMBER: US 08/289,548
; FILING DATE: 12-AUG-1994
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,940
; FILING DATE: 08-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.49964
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 2:

```



```
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-540-824-25

Query Match          7.9%; Score 80; DB 4; Length 810;
Best Local Similarity 21.4%; Pred. No. 3.7;
Matches 34; Conservative 35; Mismatches 54; Indels 36; Gaps 8;

QY 36 LVINEKSAFDVTVFG-----QRFSEILLKYDIVDGNQELRLRGFYKADSTIRKNSRI 90
Db 635 VLLNVNYESLDATISVWKSPLAFNEIILPIQOLLSAYTS---KYSDFEKPRLNKLVEKL 691
QY 91 SLREYDIKEYCNFGCAYFVLEPNRPDI-----KFD-----DERPHKRRKRSKSSOSKS 140
Db 692 KFTSHIP-----LALQNHKPVSIPTHAPKYEEFNPNPKSKSYDPDRTREINKMKA 742
QY 141 QTRNNRSOSNANAHTSKK-RKDTRRQERHIKEEQDK 178
Db 743 QLKKEK-----FTMKERIRDKARFEARQRI-EBKKE 773

RESULT 17
US-07-741-940-7
; Sequence 7, Application US/07741940
; Patent No. 5352775
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07741.940
; FILING DATE: 19920109
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.035574
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2842 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: APC

US-09-540-824-25
Query Match          7.9%; Score 80; DB 4; Length 810;
Best Local Similarity 21.4%; Pred. No. 3.7;
Matches 34; Conservative 35; Mismatches 54; Indels 36; Gaps 8;

QY 36 LVINEKSAFDVTVFG-----QRFSEILLKYDIVDGNQELRLRGFYKADSTIRKNSRI 90
Db 635 VLLNVNYESLDATISVWKSPLAFNEIILPIQOLLSAYTS---KYSDFEKPRLNKLVEKL 691
QY 91 SLREYDIKEYCNFGCAYFVLEPNRPDI-----KFD-----DERPHKRRKRSKSSOSKS 140
Db 692 KFTSHIP-----LALQNHKPVSIPTHAPKYEEFNPNPKSKSYDPDRTREINKMKA 742
QY 141 QTRNNRSOSNANAHTSKK-RKDTRRQERHIKEEQDK 178
Db 743 QLKKEK-----FTMKERIRDKARFEARQRI-EBKKE 773

RESULT 17
US-07-741-940-7
; Sequence 7, Application US/07741940
; Patent No. 5352775
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07741.940
; FILING DATE: 19920109
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.035574
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2842 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: APC

US-07-741-940-7
Query Match          7.9%; Score 79.5; DB 1; Length 2842;
Best Local Similarity 23.5%; Pred. No. 23;
Matches 28; Conservative 16; Mismatches 42; Indels 33; Gaps 3;

QY 78 YKDASTIRKNSRISLEVDYIKYCNFGCAYFVLEPNRPDIKFDDEPH--4-----126
Db 1134 VEDDRFTNYSERYSEEQHEE-----EPTNYSIKYNEKRVHVDQPIDYSLKY 1182
QY 127 -----KRRKRSKSSKSSQTSQNNRNSQSNANAHFTSKKRKDTKRRQERHIKEEQDK 177
Db 1183 ATDIPESQKQSFSEKSSGSSKSTHEMSSSEN---JSTPSSNAKRONQLHPSSAQR 1238

RESULT 18
US-08-289-548A-7
; Sequence 7, Application US/08289548A
; Patent No. 5648212
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Allegretti, LTD
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/289,548A
; FILING DATE: 12-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.46943
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2842 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: APC
; US-08-289-548A-7

US-08-289-548A-7
Query Match          7.9%; Score 79.5; DB 1; Length 2842;
Best Local Similarity 23.5%; Pred. No. 23;
Matches 28; Conservative 16; Mismatches 42; Indels 33; Gaps 3;
```


Db 33 DDPLAKSRKRKRVESDYED-----EDEVPLKRLKLSNGRAKK 71

Qy 136 QSSKSOTNRNSQSNANAHFTSKRKDKTKRQER-HIKEODKEMTSAKQ 184

Db 72 Q-VKTETKVKRKPSSANKS-KSTSKKDKTKVKRKTTVKKSATSTKVKE 119

RESULT 11

US-08-973-831-2

Sequence 2, Application US/08973831

Patent No. 5942386

GENERAL INFORMATION:

APPLICANT: Kmiec, Eric B.

APPLICANT: Gerhold, David L.

APPLICANT: Strauss, Allyson Cole

TITLE OF INVENTION: Anti-fungal Agents and Methods of

TITLE OF INVENTION: Identifying and Using the Same

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &

ADDRESSEE: No. 5942386ris

STREET: One Liberty Place, 46th floor

CITY: Philadelphia

STATE: Pennsylvania

COUNTRY: U.S.A.

ZIP: 19103

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/973.831

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/485,621

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: DeLuca, Mark

REGISTRATION NUMBER: 33,229

REFERENCE/DOCKET NUMBER: TJU-1584

TELEPHONE: 215-568-3100

TELEFAX: 215-568-3439

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 780 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-973-831-2

Query Match 8.1%; Score 82; DB 2; Length 780;

Best Local Similarity 28.2%; Pred. No. 2.2;

Matches 31; Conservative 19; Mismatches 32; Indels 28; Gaps 6;

Qy 80 DASTIRKNSRISRLP-DYIKEYCNFGCAYFVLEPNPRDIKFDDEPHKRRK---SRSKS 135

Db 33 DDPLAKSRKRKRVESDYED-----EDEVPLKRLKLSNGRAKK 71

Qy 136 QSSKSOTNRNSQSNANAHFTSKRKDKTKRQER-HIKEODKEMTSAKQ 184

Db 72 Q-VKTETKVKRKPSSANKS-KSTSKKDKTKVKRKTTVKKSATSTKVKE 119

RESULT 12

PCT-US96-09530A-2

Sequence 2, Application PC/TUS9609530A

GENERAL INFORMATION:

APPLICANT: Kmiec, Eric B.

APPLICANT: Gerhold, David L.

APPLICANT: Strauss, Allyson Cole

TITLE OF INVENTION: Anti-fungal Agents and Methods of

TITLE OF INVENTION: Identifying and Using the Same

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & Norris

STREET: One Liberty Place, 46th floor

CITY: Philadelphia

STATE: Pennsylvania

COUNTRY: U.S.A.

ZIP: 19103

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Wordperfect 6.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US96/09530A

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/485,621

FILING DATE: 07-JUN-1995

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/000,399

FILING DATE: 21-JUN-1995

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: DeLuca, Mark

REGISTRATION NUMBER: 33,229

REFERENCE/DOCKET NUMBER: TJU-1970

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-568-3100

TELEFAX: 215-568-3439

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 780 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

PCT-US96-09530A-2

Query Match 8.1%; Score 82; DB 5; Length 780;

Best Local Similarity 28.2%; Pred. No. 2.2;

Matches 31; Conservative 19; Mismatches 32; Indels 28; Gaps 6;

Qy 80 DASTIRKNSRISRLP-DYIKEYCNFGCAYFVLEPNPRDIKFDDEPHKRRK---SRSKS 135

Db 33 DDPLAKSRKRKRVESDYED-----EDEVPLKRLKLSNGRAKK 71

Qy 136 QSSKSOTNRNSQSNANAHFTSKRKDKTKRQER-HIKEODKEMTSAKQ 184

Db 72 Q-VKTETKVKRKPSSANKS-KSTSKKDKTKVKRKTTVKKSATSTKVKE 119

RESULT 13

US-09-086-912-2

Sequence 2, Application US/09086912

Patent No. 6156192

GENERAL INFORMATION:

APPLICANT: Bruce M. Spiegelman, Pere Puigserver and Zhidan Wu

TITLE OF INVENTION: PGC-1, A No. 6166192el Brown Fat PPAR(SYMBOL

TITLE OF INVENTION: 103 \f "Symbol") Coactivator

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD, LLP

STREET: 28 State Street

CITY: Boston

STATE: Massachusetts

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk


```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 236 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-299-268-28

Query Match
Best Local Similarity 8.2%; Score 82.5; DB 4; Length 236;
Matches 35; Conservative 25; Mismatches 45; Indels 39; Gaps 8;

QY 5 VTPMLN-----YKYPGPOFIHFENIVKSDIEFQVLV--INKEAFDVTVP----- 49
Db 79 VVPNHMVASIISEY---EYLKSNYKKNKINYITLDKINKYHSIDDLIFMYLHWKK 135
QY 50 -----GQFSEILLKYD-----FIVGDMGNEQLRLGRFYKDASTIRKNSRSLREDYI 97
Db 136 YNNTCAGKLFKE-LMKYDILATKYIYNDLINT-----YKEGDTISINIRLKCDDII 187
QY 98 KEYCNFGCAYFVLENNPNDRIKED 121
Db 188 K-HCKSSIGMFAILSSKIIDVDFD 210

RESULT 8
US-09-810-671-5
; Sequence 5, Application US/09810671
; Patent No. 6455291
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; FILE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; FILE OF INVENTION: THEREOF
; FILE REFERENCE: CL000758
; CURRENT APPLICATION NUMBER: US/09/810,671
; CURRENT FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 429
; TYPE: PRT
; ORGANISM: Human
US-09-810-671-5

Query Match
Best Local Similarity 8.1%; Score 82; DB 4; Length 429;
Matches 32; Conservative 23; Mismatches 39; Indels 30; Gaps 6;

QY 78 YKDASTIRKNSRISRLDYIKEYCN---FCGAYFVLENNPNDRIKFDDEPHKRR--KSR 132
Db 3 YLESRSINEKDYHSR--RYIDEYRNDYTQGC-----EPGHRQDHSR 43
QY 133 SKSQSSKSQTRNNRSQSNAN---AHFTSKKR---KDTKRQERHIKEODKEMTSAKQHL 186
Db 44 YQNHSSKSSGSGRSSYKSKHRIHSTSHRSHGKSHRRKTRTSVDEEGHLCQSGDV 103
QY 187 LFVR 190
Db 104 LSAR 107

RESULT 9
US-09-457-040B-12
; Sequence 12, Application US/09457040B
; Patent No. 6387641
; GENERAL INFORMATION:
; APPLICANT: Vertex Pharmaceuticals Incorporated
; TITLE OF INVENTION: Crystallized p38 Complexes
; FILE REFERENCE: VPI/98-14
; CURRENT APPLICATION NUMBER: US/09/457,040B
; CURRENT FILING DATE: 1999-12-08
; NUMBER OF SEQ ID NOS: 41

Query Match
Best Local Similarity 8.1%; Score 82; DB 1; Length 780;
Matches 31; Conservative 19; Mismatches 32; Indels 28; Gaps 6;

QY 80 DASTIRKNSRISRLDYIKEYCNFGCAYFVLENNPNDRIKFDDEPHKRRK---SRSKS 135
```

```
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Human
US-09-457-040B-12

Query Match
Best Local Similarity 8.1%; Score 82; DB 4; Length 484;
Matches 32; Conservative 23; Mismatches 39; Indels 30; Gaps 6;

QY 78 YKDASTIRKNSRISRLDYIKEYCN---FCGAYFVLENNPNDRIKFDDEPHKRR--KSR 132
Db 56 YLESRSINEKDYHSR--RYIDEYRNDYTQGC-----EPGHRQDHSR 96
QY 133 SKSQSSKSQTRNNRSQSNAN---AHFTSKKR---KDTKRQERHIKEODKEMTSAKQHL 186
Db 97 YQNHSSKSSGSGRSSYKSKHRIHSTSHRSHGKSHRRKTRTSVDEEGHLCQSGDV 156
QY 187 LFVR 190
Db 157 LSAR 160

RESULT 10
US-08-485-621-2
; Sequence 2, Application US/08485621
; Patent No. 5691187
; GENERAL INFORMATION:
; APPLICANT: Kmiec, Eric B.
; APPLICANT: Gerhold, David L.
; APPLICANT: Strauss, Allyson Cole
; TITLE OF INVENTION: Anti-fungal Agents and Methods of
; TITLE OF INVENTION: Identifying and Using the Same
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
; ADDRESSEE: No. 5691187ris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,621
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TJU-1584
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 780 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-485-621-2

Query Match
Best Local Similarity 8.1%; Score 82; DB 1; Length 780;
Matches 31; Conservative 19; Mismatches 32; Indels 28; Gaps 6;

QY 80 DASTIRKNSRISRLDYIKEYCNFGCAYFVLENNPNDRIKFDDEPHKRRK---SRSKS 135
```

; APPLICANT: YAN, Chunhua et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: C1000758
; CURRENT APPLICATION NUMBER: US/09/810.671
; CURRENT FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Human
US-09-810-671-2

Query Match 8.2%; Score 83; DB 4; Length 445;
Best Local Similarity 27.7%; Pred. No. 0.79;
Matches 28; Conservative 17; Mismatches 44; Indels 12; Gaps 3;
Qy 78 YKDASTIRKNSRISLEDYIKYCNFCGAYFVLEPNPNRDIKFDDEPHKRRKRSKSSQS 137
Db 20 YLEARSL--NERYDRYVDEYRNDYCEGYV-----PRHYRDIESGYRHCKSSSVRS 72
Qy 138 SKSOTRNNRSOSNANAHFTSKRKDKTKRRQERHIKEEQDK 178
Db 73 RRSPPKRR-----NRHCSHQSRKSKSHRRKRSRSEDDEE 108

RESULT 6
US-08-307-499-28
; Sequence 28, Application US/08307499
; Patent No. 5651972
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Vi uela, Eladio
; APPLICANT: Gibbs, E.P.J.
; TITLE OF INVENTION: Use of Recombinant Swine Poxvirus as a
; TITLE OF INVENTION: Live Vaccine Vector
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: U.S.A.
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/307.499
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/908,241
; FILING DATE: 1-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/908,630
; FILING DATE: 29-JUN-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/342,212
; FILING DATE: 21-APR-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UF35.1.FWCC1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800

; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 236 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-307-499-28

Query Match 8.2%; Score 82.5; DB 1; Length 236;
Best Local Similarity 24.3%; Pred. No. 0.38;
Matches 35; Conservative 25; Mismatches 45; Indels 39; Gaps 8;
Qy 5 VTPEMLN-----YNKYPGQFIHFENIVKSDDIEFOLV--INERSAFDVTVF----- 49
Db 79 VVPHMNVASIIYSEY---EYLLKSNVKNKKINTYLDKINKYHSIDDIIFMYLHWKK 135
Qy 50 -----GQRSEILLKYD-----FIVGDMGNEQLRLRGFYKDASTIRKNSRISLEDYI 97
Db 136 YNNTCACGKLFKE-LMKYDILATKYIYNDIINT-----YKEGDTISINIRLCKCKDDII 187
Qy 98 KEYCNFCGAYFVLEPNPNRDIKFD 121
Db 188 K-HCKSSIGMFAILLSSKITIDVDFD 210

RESULT 7
US-09-299-268-28
; Sequence 28, Application US/09299268
; Patent No. 6217882
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Vi uela, Eladio
; APPLICANT: Gibbs, E.P.J.
; TITLE OF INVENTION: Use of Recombinant Swine Poxvirus as a
; TITLE OF INVENTION: Live Vaccine Vector
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: U.S.A.
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/299.268
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/901,127
; FILING DATE:
; APPLICATION NUMBER: US 07/908,241
; FILING DATE: 1-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/908,630
; FILING DATE: 29-JUN-1992
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/342,212
; FILING DATE: 21-APR-1992
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UF35.1.FWCC1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 28:

```
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/667,276A
; FILING DATE: 11-MAR-1991
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 377,5351P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 414 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-667-276A-4

Query Match 8.5%; Score 86; DB 1; Length 414;
Best Local Similarity 26.3%; Pred. No. 0.34;
Matches 18; Conservative 20; Mismatches 28; Indels 2; Gaps 1;

QY 117 DIKFDDEPHKRRKSKSQSSQTSOTRNNRSQSNANAHFTSKRRKDTKRRQERHIKEEQD 176
Db 94 DSSSDEE--EEEEKETKKEESKSSSSSSSSSSSSSESEKESNDKKRKSDAEDEED 151
QY 177 KEMTSKQ 184
Db 152 EESSNKKQ 159

RESULT 4
US-09-810-671-4
; Sequence 4, Application US/09810671
; Patent No. 6455291
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; FILE REFERENCE: CL000758
; CURRENT APPLICATION NUMBER: US/09/810,671
; CURRENT FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 427
; TYPE: PRT
; ORGANISM: Human
; US-09-810-671-4

Query Match 8.2%; Score 83; DB 4; Length 427;
Best Local Similarity 27.7%; Pred. No. 0.75;
Matches 28; Conservative 17; Mismatches 44; Indels 12; Gaps 3;

QY 78 YKDASTRKNSRISRLEDYIKYCNFCAYFVLENPNRDIKFDDEPHKRRKSKSQS 137
Db 2 YLEARSLL--NERDYDRRYVDEYRNDYCEGYV-----PRHYHRDIESGYRHCSSKSVRS 54
QY 138 SKSOTRNNRSQSNANAHFTSKRRKDTKRRQERHIKEEQDKE 178
Db 55 RRSSPKRR-----NRHCSSHQSKSKSHRRKRSKSIEDDEE 90

RESULT 5
US-09-810-671-2
; Sequence 2, Application US/09810671
; Patent No. 6455291
; GENERAL INFORMATION:
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2003, 09:51:52 ; Search time 15 seconds
(without alignments)
376.613 Million cell updates/sec

Title: US-09-868-352-23
Perfect score: 1009
Sequence: 1 MRKEVTEMLNKNYPGPQF.....EEQDKEMTSKQHLLEVRKN 192

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	254.5	25.2	132	4	US-09-134-001C-4894
2	87	8.6	732	4	Sequence 4894, Ap
3	86	8.5	414	1	Sequence 8, Appli
4	83	8.2	427	4	Sequence 4, Appli
5	83	8.2	445	4	Sequence 2, Appli
6	82.5	8.2	236	1	Sequence 28, Appli
7	82.5	8.2	236	4	Sequence 28, Appli
8	82	8.1	429	4	Sequence 5, Appli
9	82	8.1	484	4	Sequence 12, Appli
10	82	8.1	780	1	Sequence 2, Appli
11	82	8.1	780	2	Sequence 2, Appli
12	82	8.1	780	5	PCT-US96-09530A-2
13	82	8.1	797	4	Sequence 2, Appli
14	82	8.1	797	4	Sequence 2, Appli
15	80	7.9	262	4	Sequence 2, Appli
16	80	7.9	810	4	Sequence 5532, Ap
17	79.5	7.9	2842	1	Sequence 25, Appli
18	79.5	7.9	2842	1	Sequence 7, Appli
19	79.5	7.9	2842	1	Sequence 7, Appli
20	79.5	7.9	2842	1	Sequence 7, Appli
21	79.5	7.9	2843	1	Sequence 2, Appli
22	79.5	7.9	2843	1	Sequence 2, Appli
23	79.5	7.9	2843	1	Sequence 2, Appli
24	79.5	7.9	2843	1	Sequence 2, Appli
25	79.5	7.9	2843	1	Sequence 2, Appli
26	79.5	7.9	2843	2	Sequence 2, Appli
27	79.5	7.9	2843	3	Sequence 2, Appli

28 79.5 7.9 2843 3 US-08-450-582-7 Sequence 7, Appli
29 79.5 7.9 2843 4 US-08-449-731-2 Sequence 2, Appli
30 79.5 7.9 2973 2 US-08-821-355A-7 Sequence 7, Appli
31 79.5 7.9 2973 2 US-09-003-687A-7 Sequence 7, Appli
32 79.5 7.9 2973 4 US-09-136-605-7 Sequence 7, Appli
33 79 7.8 475 2 US-08-698-407-4 Sequence 4, Appli
34 79 7.8 475 3 US-09-195-855-4 Sequence 5, Appli
35 79 7.8 798 4 US-09-203-453-5 Sequence 5, Appli
36 78.5 7.8 1182 4 US-09-041-886-21 Sequence 21, Appli
37 78.5 7.8 2265 2 US-08-149-097D-36 Sequence 36, Appli
38 78.5 7.8 2509 2 US-08-149-097D-35 Sequence 35, Appli
39 77.5 7.7 564 4 US-09-134-001C-3930 Sequence 3930, Ap
40 77.5 7.7 1048 3 US-09-356-952-5 Sequence 5, Appli
41 77.5 7.7 1898 1 US-08-056-200-94 Sequence 94, Appli
42 77.5 7.7 1898 2 US-08-800-644-94 Sequence 94, Appli
43 76.5 7.6 265 4 US-09-149-476-497 Sequence 497, App
44 76.5 7.6 216 1 US-08-185-424B-1 Sequence 1, Appli
45 76.5 7.6 432 2 US-08-933-750C-47 Sequence 47, Appli

ALIGNMENTS

RESULT 1
US-09-134-001C-4894
; Sequence 4894, Application US/09134001C
; Patent No. 6300370
; GENERAL INFORMATION:
; APPLICANT: L'inn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4894
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4894

Query Match 25.2%; Score 254.5; DB 4; Length 132;
Best Local Similarity 45.6%; Pred. No. 4.9e-20;
Matches 47; Conservative 24; Mismatches 29; Indels 3; Gaps 2;
Qy 24 ENIVKSDIEFQLVINEKSAFDVTVEGOFSEILLKYDFIVGDWGNQOLRLRGFYKDAST 83
Db 2 ENMIVDQOQYFELIEYRECFDEEIFSARYSDILDKYDVVGDIYDQLRLAGFYRDSNK 61
Qy 84 IRK-NSRISRLIEDYIKYCNFCGAYFVLENNPRDI--KFDDE 123
Db 62 KAEIKRESSIQDYILEYCNFCGAYFVVRRLSPNEFIEIDDK 104

RESULT 2
US-08-914-999-8
; Sequence 8, Application US/08914999
; Patent No. 6346406
; GENERAL INFORMATION:
; APPLICANT: Ryazanov, Alexey G.
; APPLICANT: Hait, William N.
; APPLICANT: Pavur, Karen S.
; TITLE OF INVENTION: ELONGATION FACTOR-2 KINASE (EF-2 KINASE)
; TITLE OF INVENTION: ELONGATION FACTOR-2 KINASE (EF-2 KINASE)
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESS: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; FLOOR

```

C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001
C:Accession: JC7720
R:Choi, S.C.; Chang, J.Y.; Han, J.K.
Biochem. Biophys. Res. Commun. 285, 1338-1343, 2001
A:title: A novel Xenopus acetyltransferase with a dynamic expression in early develop
A:Reference number: JC7720; MUID:21371756; PMID:11478804.
A:Contents: Eggs
A:Accession: JC7720
A:Molecule type: mRNA
A:Residues: 1-846 <CHO>
A:Cross-references: GB:AF247679
C:Comment: This enzyme, a novel acetyltransferase containing two tetratricopeptide r
s.
C:Genetics:
A:Gene: Xat-1
C:Keywords: acyltransferase

Query Match      8.8%; Score 89; DB 2; Length 846;
Best Local Similarity 22.4%; Pred. No. 19;
Matches 41; Conservative 27; Mismatches 69; Indels 46; Gaps 7;

QY 13 NKYPGPQFIHFENIVKSDDIEFOLVINEKSAFDVTVFGORFSEI-----LLKYDFIVGDM 67
Db 481 NKYG-----EALKKCHEIERHFVELTDQDFHTYCMRKITLRSYADLLKLE----- 527

QY 68 GNEQLRLGKYQDASTIRKNSIRSLREYDIKEYCNFGCAYFVLEPNPNRDIKFDDEPRHK 127
Db 528 --DVLROHPYFKAARI-----AIEIVLKLH-----DNP-----LTDENKEH 562

QY 128 RRKRSKSSQSSKSQTRNNRSQSNANAHFTSKKKRDKYRQERH---IKEEQDKEMTSAKO 184
Db 563 EADTANMSDKELKLRNKRQAOKKAQLEBEKKNAEKEKQORNQKKKKEDDDEEIGGPKE 622

QY 185 HLL 187
Db 623 ELI 625

RESULT 40
T31095
vitellogenin precursor - Oreochromis aureus
C:Species: Oreochromis aureus
C>Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-May-2000
C:Accession: T31095
R:Lim, E.H.; Lam, T.J.; Ding, J.L.
submitted to the EMBL Data Library, August 1997
A:Description: Cloning of full-length Oreochromis aureus vitellogenin cDNA and its de
A:Reference number: Z20978
A:Accession: T31095
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1788 <LIM>
A:Cross-references: EMBL:AF017250; NID:g4102880; PID:g4102881; PIDN:AAD01615.1
C:Superfamily: vitellogenin

Query Match      8.8%; Score 89; DB 2; Length 1788;
Best Local Similarity 31.3%; Pred. No. 44;
Matches 21; Conservative 16; Mismatches 28; Indels 2; Gaps 1;

QY 128 RRKRSKSSQSSKSQTRNNRSQSNANAHFTSKKKRDKYRQERHIKEEQDKEMTSAKQHL 187
Db 1219 KRKRSRSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSKSKRSKSSKSSSSSSRRSAQSKQL 1278

QY 188 FV--RKN 192
Db 1279 ALKFRKN 1285

Search completed: February 26, 2003, 09:54:42
Job time : 22 secs

```


Db 36 GYGFVEFEDSRDADDAVVELNGKELCGEHVIVEHARGPRDRDGYSGYSGSGGGYSSRR 95
Qy 74 LRGFYKDASTIRKNSRI-----SR-----LEDYI-----KEYCNFGCAYF-- 108
Db 96 TSGRDYKGPVTEYTRVLIVENUSRCSCWDLKDFMRQAGEVTVYADAHKERTWEGVIEFRS 155
Qy 109 -----VLE-----NPNPRDIKFDDEP-----HKRRKRSRKS-OSSKSO 141
Db 156 YSDMKRALDKLDGTEINGENIRLIEDKPTSHRRSYSGSRSRSSRSRSSRSR 215
Qy 142 TRN-NRSQSNANAHFTSKRKDTKRQERHIKEEQDKEMTSAKQH 185
Db 216 SRISKRSRSGSRSGRSRSGKSGKRSKSKPKSGKSGSHS 260
RESULT 27
E70013
Hypothetical protein yukC - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
C:Accession: E70013
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertoni, C.; Bron, S.; Brouillette, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chao, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E. Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallier, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinols, A.; Laubert, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron, Akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K. A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: E70013
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-451 <KUN>
A:Cross-references: GB:299120; GB:AL009126; NID:g2635613; PIDN:CAB15177.1; PID:ell84366;
A:Experimental source: strain 168
C:Genetics:
A:Gene: yukC
Query Match 9.1%; Score 91.5; DB 2; Length 451;
Best Local Similarity 21.9%; Pred. No. 6.1;
Matches 44; Conservative 31; Mismatches 71; Indels 55; Gaps 9;
Qy 4 EVTPMLNKNYKPGQFIHFENI--VKSDIEFQVLVINEKSAFDVTVFGQRFSEILLKYD 61
Db 275 ESLPSVQYL--ATSYVEVENLGSAAKTKNIENNVLTQSDP-----QHFLYWI----- 321
Qy 62 FTVGQWNEQLRLRGFYKDASTIRKNSRISRLDYIKEYCNFGCAYFVL-----EN 112
Db 322 ---DYG-----RGEYKATSI---GRKLENDYI-----YFALAKYKQLLSED 359
Qy 113 PNPRDI-----KFDDERPHKRRKRSKSSQSKSQSTRNRSQSNANAHFTSKRKDT 163
Db 360 TNDEIQKELDSVNSELEKAQKERQENQSNSETSLVDTSEBQTQDEKQAEKAAEEK 419
Qy 164 KRQERHIKEEQDKEMTSAKQ 184
Db 420 AAAEEKAKKEQKEDEKKE 440
RESULT 28
A47328
natural killer cell tumor-recognition protein - human
N:Alternate names: cyclophilin-related NK-tumor recognition protein; natural killer-tum
C:Species: Homo sapiens (man)

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 24-Sep-1999
C:Accession: A47328
R:Anderson, S.K.; Gallinger, S.; Roder, J.; Frey, J.; Young, H.A.; Ortaldo, J.R. Proc. Natl. Acad. Sci. U.S.A. 90, 542-546, 1993
A:Title: A cyclophilin-related protein involved in the function of natural killer cell
A:Reference number: A47328; MUID:9313824; PMID:8421688
A:Accession: A47328
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1403 <AND>
A:Cross-references: L:1403 <AND> PIDN:AAA35734.1; PID:g181252
A:Experimental source: NK killer cells from adult blood
A:Note: sequence extracted from NCBI backbone (NCBIN:122798, NCBIP:122800)
C:Genetics:
A:Gene: GDB:NKTF
A:Cross-references: GDB:137171; OMIM:161565
A:Map position: 3p23-3p21
C:Superfamily: natural killer cell tumor-recognition protein; cyclophilin homology
C:Keywords: alternative splicing; lymphocyte
F:60-230/Domain: cyclophilin homology <CYP>
Query Match 9.1%; Score 91.5; DB 1; Length 1403;
Best Local Similarity 28.1%; Pred. No. 21;
Matches 32; Conservative 14; Mismatches 41; Indels 27; Gaps 4;
Qy 78 YKD---ASTIRKNSRISRLDYIKEYCNFGCAYFVLENPNDRIKFDDEPHKRRKRSK 134
Db 735 YSDRSKSSSRSSRSRSGYSTR-----SRSSSSSSRSR 774
Qy 135 SQSSKSKQTRN---NRQSNANAHFTSKRKDTKRQERHIKEEQDKEMTSAKQH 185
Db 775 SPSSKSHSRNYSYDHSQSSSSSYTIS-SDGRRKRRRLSSGKNSVSHKKH 827
RESULT 29
S57157
Hypothetical protein YJ134c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein J2120
C:Species: Saccharomyces cerevisiae
C:Date: 23-Aug-1995 #sequence_revision 08-Sep-1995 #text_change 19-Apr-2002
C:Accession: S57157
R:Rose, M.; Koetter, P.; Entian, K.D.
A:Reference number: S56848
A:Accession: S57157
A:Molecule type: DNA
A:Residues: 1-737 <ROS>
A:Cross-references: EMBL:Z49634; NID:g1015869; PIDN:CAA89665.1; PID:g1015870; GSPDB:G
A:Gene: SGD:SGM1; MIPS:YJ134C
A:Cross-references: SGD:S0003895
A:Map position: 10R
Query Match 9.0%; Score 90.5; DB 2; Length 707;
Best Local Similarity 22.2%; Pred. No. 12;
Matches 41; Conservative 33; Mismatches 52; Indels 59; Gaps 8;
Qy 3 KEVTEMLNKNYKPGQFIHFENIYKSDIEFQVLVINEKSAFDVTVFGQRFSEILLKYDF 62
Db 127 KEISFEYLRNKK-----QIDLTNEL--NRKSQIETT--DSSFFKLKEKDD 169
Qy 63 IVGDKNGEQLRLRGFYKDASTIRKNSRISRLDYIKEYCNFGCAYFVLENPNDRIKFD 122
Db 170 LI---DQLRKEGAKLAETELRQSNQIKALTKVK-----DLEYE- 205
Qy 123 ERPHARRKRSKSSQSKSQSTRNRSQS-----NANAHFTSK-----RKDKRRQERHI 171
Db 206 -----VSELNDSSAQSVENNELQSLYHNITQGLAEATNKLKADQKESLETLEKNI 258
Qy 172 KEEQI 176
Db 259 KEKDI 263

A:Status: preliminary; translated from GB/EMBL/DDBBJ
A:Molecule type: DNA
A:residues: 1-1650 <LAW>
A:Cross-references: EMBL:Z98547; NID:e1325376; PID:
C:Genetics:
A:Map position: 3
A:Introns: 1597/3; 1625/3
A:Note: C0385C

	Query Match	9.4%	Score 94.5;	DB 2;	Length 1650;
	Best Local Similarity	21.5%;	Pred.	No. 15;	
	Matches 43; Conservative	42;	Mismatches	74;	Indels 41; Gaps 10;
Qy	3 KEVT-PEMLNKNYPGPGQFHFN----	IWKSDDIEFQLVINKSAFDVTFVGORFSEI	56	:	: :
Dd	406 KEINKEEYSNKNEY-----VFHKNDSSIKKKNSSECLEOQRKKKYIIIOQ----	456	:	: :	:
Qy	57 LLKYDYFTVDWGNBOL-----RLRGPY-----KDASTIRKKSRSRLEDYIKEYCNF	103	:	: : :	:
Dd	457 --RYNFNDRONNNAYIKDDTHHKEKGYYLNVIVQSERYKYGSGSNKMDEMEIYNQHTNDF	514	:	: :	:
Qy	104 GCAYFVLLENPNDRIDEPHPKRKRSRKSQ---	SSKSQTNRNRSQSANAAHFTSKK	159	:	: :
Dd	515 N----INENLNK--IYFDTEGYDPERKKKLDDHHITTOQEYRNKNNIDLTKDHHLNDKE	569	:	: : :	:
Qy	160 RKDTKRRROERHIKEEQDKEM	179	:	: : :	:
Dd	570 TKE--KRNEIEIEEEKNKNI	587	:	: : :	:

RESULT 24
T18473
hypothetical protein C0435w - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T18473
R:Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, November 1998
A:Reference number: Z18937
A:Accession: T18473
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1294 <LAW>
A:Cross-references: EMBL:AL008970; NID:e1407852; PID:e1332567; PIDN:CAA15616.1
C:Genetics:
A:Map position: 3
A:Note: C0435w

Query Match	9.2%	Score 93:	DB 2;	Length 1294;
Best Local Similarity	21.4%;	Pred. No. 15;		
Matches	39;	Conservative	31;	Mismatches 78; Indels 34; Gaps 16;
QY	3	KEVTPMLNKNKYPGQPIHFHFNIVKSDDIEFQLV-----	IN EKSAFDVTVVF-----	49
Db	609	KEGLCEKLKHMIV----	Y Y Y E Y V K K S E K K Y F I A D D T F V N V K N L D V N L N L N T C S H	664
QY	50	--GGRFSEILKLYDFIVGDWNGEOLRLRGFYKDA-----	STIRKNSRISRLDYIKE	99
Db	665	SKKYMYDKYIKSYDFVK---	E N E A L F L Q N F P K T L F L Y S Y L K D T F A K T I Q T L K K Y D Y V P K	721
QY	100	YC N F G C A V F V L E N P R D I K E D D E R P H K R R K S R S K S O S K S Q T R N N R S O S N A N A H F T S K K	159	
Db	722	Y C Q G G -----	I L S K H K N N D S D D D H H V G N K Q N N D S T H Q D I E K N Q V N V I N N N N N N N N K	777
QY	160	RK	161	
Db	778	AK	779	

RESULT 25
B47328
natural killer cell tumor-recognition protein - mouse
N:Alternate names: cyclophilin-related NK-tumor recognition protein; natural killer-tumor

C/Species: Mus musculus (house mouse)
C/Date: 17-Jul-1994 #sequence_revision 17-Jul-1994 #text_change 05-Nov-1999
C/Accession: B47328; I77662
R/Anderson, S.K.; Gallinger, S.; Roder, J.; Frey, J.; Young, H.A.; Ortaldo, J.R.
Proc. Natl. Acad. Sci. U.S.A. 90, 543-546, 1993
A/Title: A cyclophilin-related protein involved in the function of natural killer cell
A/Reference number: A47328; MUID:93I33824; PMID:8421688
A/Accession: B47328
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-1507 <AND>
A/Cross-references: GB:L04289; NID:g192866
A/Note: authors translated the codon AGT for residue 972 as Arg
R/Rinfret, A.; Anderson, S.K.
Mol. Immunol. 30, 1307-1313, 1993
A/Title: IL-2 regulates the expression of the NK-TR gene via an alternate RNA splicing
A/Reference number: I57820; MUID:94019422; PMID:8413330
A/Accession: I77662
A/Status: preliminary; translated from GB/EMBL/DBD
A/Molecule type: mRNA
A/Residues: 235-237;263-294 <RIN>
A/Cross-references: GB:S65598; NID:g425701; PIDN:AAB28500.1; PID:g425702
C/Genetics:
A/Gene: NK-TR
C/Superfamily: natural killer cell tumor-recognition protein; cyclophilin homology
C/Keywords: alternative splicing; lymphocyte
F:60-230/Domain: cyclophilin homology <CYP>

Query Match	9.2%;	Score 93;	DB 2;	Length 1507;	
Best Local Similarity	20.2%;	Pred. No. 17;			
Matches	38;	Conservative ~ 35;	Mismatches 61;	Indels 54;	Gaps 6;
QY	50	GORPSEILLKY----	DFIVGDWGNELRLRGFYKDKASTIRKNSRISRLPEDIYKEYCNEG	104	
Db					
QY	115	GSTRVVKVNMIOGGDFSEGNKGGSIIYGGYFKDENFILKHDRAFLLSMANRGKHTNG	174		
Db					
QY	105	CAYFVLENPNP-----	RDIKED-DERPH-----	126	
Db					
QY	175	SQFTITTKPAPHLDGVHVFGLVIGSGFEVIEQIENLATDAASRPYADVVRVDCGLVATKL	234		
Db					
QY	127	-----KKRSKRSKSQSKSTQRNNRSOSNAN----	AHFTSKRKDKTKRQERHIKEED	176	
Db					
QY	235	TKDVFENKRRKPTCEGSDSSRSSSSSSSSSESEVERETIRRRRHHRRRPVVRHAKRRR-	293		
Db					
QY	177	KEMTSAKQ	184		
Db					
QY	294	KEMSSSEE	301		
Db					

RESULT 26
S59043
spilling factor Srp55 - human
C:Species: Homo sapiens (man)
C:Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 28-May-1999
C:Accession: S59043
R:Sreaton, G.R.; Caceres, J.F.; Mayeda, A.; Bell, M.V.; Plebanski, M.; Jackson, D.G.
EMBO J. 14, 4336-4349, 1995
A:Title: Identification and characterization of three members of the human SR family
A:Reference number: S59042; MUID: 96016206; PMID: 7556075
A:Accession: S59043
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-344 <SCR>
A:Cross-references: GB:U30883; NID:g1049087; PIDN:AAA93073.1; PID:g1049088
C:Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleoprotein
F:3-62/Domain: ribonucleoprotein repeat homology <RRM1>
F:111-173/Domain: ribonucleoprotein repeat homology <RRM2>

```
'Query Match          9.1%; Score 92; DB 2; Length 344;
Best Local Similarity 23.1%; Pred. No. 4.2;
Matches 52; Conservative 34; Mismatches 83; Indels 56; Gaps 9;

Qv 17 GPQFIHFENIVKSDIEFOLVINEKSAFDVTV---FGQRFSEILLKYDFIVGDWGNQOLR 73
```


Query Match 10.9%; Score 109.5; DB 2; Length 1264;
Best Local Similarity 24.8%; Pred. No. 0.68;
Matches 55; Conservative 31; Mismatches 67; Indels 69; Gaps 12;

QY 3 KEVTP--EMLNKKY--PG-POFIHFENIVKSDDLIEFQVLVINEKSAFADVTFVGFSEI 56
DB 140 QOSTPQDDMEVNSVGNPLQPTQYORRIVIREKVOFNATINPES-----REI 188
QY 57 LL-----KY-----DFIYDGMNEQLRLRGFYKDASTI-----RKN 87
DB 189 VPAHTKYMLIEAASNNKSEIV-DQOQ-----IYDRSTFTKHLIKCFEKITLQAS 240
QY 88 SRTSLREDIYKEVNCGCAVFLVLENPNRD-IKFDDERPHKRRKRSKSSQ-----137
DB 241 SKGAMPCKVPEY----LAMYGLTMEMPDKMLYKEDEPVVARRSNSANVSPSEKKNR 296
QY 138 -SKSOTRNNRSQSNANAHFTSKRKDKTKRQERHIKEEQDK 178
DB 297 QSKSGKSNSTNDASNKKETKRRKKPTEVNDSENSEEDKK 338

RESULT 10
T27245
hypothetical protein Y57G11C.9a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T27245
R:McMurray, A.
submitted to the EMBL Data Library, September 1997
A:Reference number: Z20330
A:Accession: T27245
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-553 <WIL>
A:Cross-references: EMBL:Z99281; PIDN:CAB54457.1; GSPDB:GN00022; CESP:Y57G11C.9a
A:Experimental source: clone Y57G11C
C:Genetics:
A:Gene: CESP:Y57G11C.9a
A:Map position: 4
A:Introns: 67/2; 109/2; 382/1; 418/2; 447/1; 496/3

Query Match 10.4%; Score 105; DB 2; Length 553;
Best Local Similarity 23.4%; Pred. No. 0.64;
Matches 46; Conservative 24; Mismatches 93; Indels 34; Gaps 5;

QY 11 NYNKPQPFQFIHFENIVKSDDI-----EFQLVINE---KSAFDV 46
DB 22 NYNNNNHHHHYNNLPINETQVPSKHTFIRGLHGLDSTEEIKYIGEKVKISFDF 81
QY 47 TVFGQRFSEILLKYDFIVDGMNEQLRLRGFYKDASTIRKNSRISRLIEDYIKEYCNPFGCA 106
DB 82 VKVAQDKSKIFVAVRFENRDEAKEFMET---YSDREFMGCRCDLSPFRD-IRRYCAYORA 137
QY 107 YFVLEPNPRDIKFDDEPHKR-----RKSRKSSQSSKSTQTRNNRSQSNANAHFTSKKR 160
DB 138 KQVRSNQRRRRRSDQESSKRSASPPVRRGRSKRSRSPSRSSRSRSRSVSRKSP 197
QY 161 KDKRQERHIKEEQDK 177
DB 198 ASRNGRVSRHRSRSONR 214

RESULT 12
E71620
hypothetical protein PFB0225c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C:Accession: E71620
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.
; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743; PMID:9804551
A:Accession: E71620
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-508 <GAR>
A:Cross-references: GB:AE001381; GB:AE001362; NID:g3845124; PIDN:AAC71834.1; PID:g384
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PFB0225c

Query Match 10.0%; Score 101; DB 2; Length 508;
Best Local Similarity 20.9%; Pred. No. 1.2;
Matches 41; Conservative 38; Mismatches 61; Indels 56; Gaps 9;

QY 5 VTEMLNKNYKPGQFIHFENIV-----KSDDIE---FQLVINEKSAFADVTFVFG 50
DB 264 ITKE-CNVNKYD-----DNIIDHKQHKREKKEKKSINNMINSHIYEKEQSDIC---312
QY 51 QRFSEILLKYDFIVDGMNEQLRLRGFYKDASTIRKNSRISRLIEDY--IKEYCNFGCAYF 108
DB 313 -----NVLEENKEEYNNLQKDVITNCNNDKV-KLEEYHHEKELNN-----353
QY 109 VLENPNPRDIKFDDEPHKRRKRSKSSQSSKSTQTRNNRSQSNANAHFTSKRDKTKRROE 168
DB 354 -VOIINDMDIKKNEAKKEKNKKKQKNKKNEKNKKK-----EKENKKKEKE 403
QY 169 RHKEQDKEMTSAKQ 184
DB 404 KKKKEKSKKKEKE 419

RESULT 13

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-659 <WIL>
A:Cross-references: EMBL:Z99281; PIDN:CAB54458.1; GSPDB:GN00022; CESP:Y57G11C.9b
A:Experimental source: clone Y57G11C
C:Genetics:
A:Gene: CESP:Y57G11C.9b
A:Map position: 4
A:Introns: 67/2; 109/2; 382/1; 418/2; 524/2; 553/1; 602/3

Query Match 10.4%; Score 105; DB 2; Length 659;
Best Local Similarity 23.4%; Pred. No. 0.77;
Matches 46; Conservative 24; Mismatches 93; Indels 34; Gaps 5;

QY 11 NYNKPQPFQFIHFENIVKSDDI-----EFQLVINE---KSAFDV 46
DB 22 NYNNNNHHHHYNNLPINETQVPSKHTFIRGLHGLDSTEEIKYIGEKVKISFDF 81
QY 47 TVFGQRFSEILLKYDFIVDGMNEQLRLRGFYKDASTIRKNSRISRLIEDYIKEYCNPFGCA 106
DB 82 VKVAQDKSKIFVAVRFENRDEAKEFMET---YSDREFMGCRCDLSPFRD-IRRYCAYORA 137
QY 107 YFVLEPNPRDIKFDDEPHKR-----RKSRKSSQSSKSTQTRNNRSQSNANAHFTSKKR 160
DB 138 KQVRSNQRRRRRSDQESSKRSASPPVRRGRSKRSRSPSRSSRSRSVSRKSP 197
QY 161 KDKRQERHIKEEQDK 177
DB 198 ASRNGRVSRHRSRSONR 214

RESULT 12
E71620
hypothetical protein PFB0225c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C:Accession: E71620
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.
; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743; PMID:9804551
A:Accession: E71620
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-508 <GAR>
A:Cross-references: GB:AE001381; GB:AE001362; NID:g3845124; PIDN:AAC71834.1; PID:g384
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PFB0225c

Query Match 10.0%; Score 101; DB 2; Length 508;
Best Local Similarity 20.9%; Pred. No. 1.2;
Matches 41; Conservative 38; Mismatches 61; Indels 56; Gaps 9;

QY 5 VTEMLNKNYKPGQFIHFENIV-----KSDDIE---FQLVINEKSAFADVTFVFG 50
DB 264 ITKE-CNVNKYD-----DNIIDHKQHKREKKEKKSINNMINSHIYEKEQSDIC---312
QY 51 QRFSEILLKYDFIVDGMNEQLRLRGFYKDASTIRKNSRISRLIEDY--IKEYCNFGCAYF 108
DB 313 -----NVLEENKEEYNNLQKDVITNCNNDKV-KLEEYHHEKELNN-----353
QY 109 VLENPNPRDIKFDDEPHKRRKRSKSSQSSKSTQTRNNRSQSNANAHFTSKRDKTKRROE 168
DB 354 -VOIINDMDIKKNEAKKEKNKKKQKNKKNEKNKKK-----EKENKKKEKE 403
QY 169 RHKEQDKEMTSAKQ 184
DB 404 KKKKEKSKKKEKE 419

RESULT 13

AH1744
B: subtilis yutD protein homolog lin2501 [imported] - Listeria innocua (strain Clip11262)
C:Species: Listeria innocua
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 17-May-2002
C:Accession: AH1744
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kretz, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mak, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AH1744
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-91 <GLA>
A:Cross-references: GB:AL592022; PIDN:CAC97728.1; PID:g16415023; GSPDB:GN00178
A:Experimental source: strain Clip11262
C:Genetics:
A:Gene: lin2501
C:Superfamily: Bacillus subtilis hypothetical protein yutD
Query Match 23.7%; Score 239.5; DB 2; Length 91;
Best Local Similarity 51.2%; Pred. No. 1.5e-12;
Matches 43; Conservative 22; Mismatches 14; Indels 5; Gaps 2;
Qy 31 DIEFQLVINEKSAFVTVFGQRFSEILLKYDFIVGDWGNQQLRLRGFYKDASTIRK---N 87
Db 7 DLNVEITNYDAFDEEKLNRFSIDILGRDYIVGDWGYDQLRLKGFEDDN--RKAAYD 64
Qy 88 SRISLEDYIKYCNFGCAYFVLE 111
Db 65 NKISTLKEYIYECNFGCAYFVIR 88
RESULT 7
A84079
hypothetical protein BH3433 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 17-May-2002
C:Accession: A84079
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira, Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: A84079
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-93 <STO>
A:Cross-references: GB:AP001518; GB:BA000004; NID:g10175792; PIDN:BA07152.1; GSPDB:GN00178
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH3433
C:Superfamily: Bacillus subtilis hypothetical protein yutD
Query Match 23.2%; Score 234.5; DB 2; Length 93;
Best Local Similarity 49.4%; Pred. No. 3.9e-12;
Matches 44; Conservative 21; Mismatches 19; Indels 5; Gaps 2;
Qy 26 IVKSDIEFQLVINEKSAFVTVFGQRFSEILLKYDFIVGDWGNQQLRLRGFYKD---AS 82
Db 1 MVRISNQFVIENVKDGNNEAFVRYSDVLNKYDYIVGDWGYDQLRLKGFEDDNKKS 60
Qy 83 TIRKNSRISLEDYIKYCNFGCAYFVLE 111
Db 61 TYDK--KISTLDYIYECNFGCAYFVVK 87
RESULT 8
F70023
hypothetical protein yutD - Bacillus subtilis
C:Species: Bacillus subtilis

C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
C:Accession: F70023
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber, C.; Bron, S.; Brouillet, S.; Bruchli, C.V.; Calowell, B.; Capuano, V.; Carter, N.M.; A.; Ehrlich, S.E.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal, lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino, A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau, Y., M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete, Rieger, M.; Rivolta, C.; Rocha, E.; Roche, M.; Rose, M.; Sadiq, Y.; Sato, T.; Scanl, A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se, akeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: F7023
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-112 <RUN>
A:Cross-references: GB:299120; GB:AL009126; NID:g2635613; PIDN:CAB15221.1; PID:g26357 A:Experimental source: strain 168
C:Genetics:
A:Gene: yutD
C:Superfamily: Bacillus subtilis hypothetical protein yutD
Query Match 23.0%; Score 232.5; DB 2; Length 102;
Best Local Similarity 51.7%; Pred. No. 6.2e-12;
Matches 46; Conservative 16; Mismatches 26; Indels 1; Gaps 1;
Qy 24 ENIVK:DDIEFQLVINEKSAFVTVFGQRFSEILLKYDFIVGDWGNQQLRLRGFYKDA- 82
Db 10 EIMIL:QNAEFELVHNFKDFGNEAFKARYSDILNKYDYIVGDWGYDQLRLKGFDDQNG 69
Qy 83 TIRKNSRISLEDYIKYCNFGCAYFVLE 111
Db 70 KATFE:KISTLDYIYECNFGCAYFVLE 98
RESULT 9
S64146
probable membrane protein YGL133w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein G2842
C:Species: Saccharomyces cerevisiae
C:Date: 17-May-1996 #sequence_revision 17-May-1996 #text_change 19-Apr-2002
C:Accession: S64146; S71739
R:Escaribano, V.; Eraso, P.; Portillo, F.; Mazon, M.J.
submitted to the Protein Sequence Database, May 1996
A:Reference number: S64144
A:Accession: S64146
A:Molecule type: DNA
A:Residues: 1-1364 <ESC>
A:Cross-references: EMBL:Z72655; NID:g1322701; PID:e243361; PID:g1322702; MIPS:YGL133 A:Experimental source: strain S288C
R:Escaribano, V.; Eraso, P.; Portillo, F.; Mazon, M.J.
Yeast 12, 887-892, 1996
A:Title: Sequence analysis of a 14.6 kb DNA fragment of Saccharomyces cerevisiae chro A:Reference number: S71733; MUID:96437978; PMID:8840506
A:Accession: S71739
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1264 <ESW>
A:Cross-references: EMBL:X92670
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1995 C:Genetics:
A:Gene: SGD:ITC1
A:Cross-references: SGD:S0003101
A:Map position: 7L
C:Keywords: transmembrane protein
F:549-565/Domain: transmembrane #status predicted <TM>

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2003, 09:51:32 ; Search time 17 Seconds
(without alignments)
1085.753 Million cell updates/sec

Title: US-09-868-352-23
Perfect score: 1009
Sequence: 1 MRKEVTPMLNKNYPGPQF.....EEQDKEMTSKQHLLEVRKN 192

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	452.5	44.8	176	C97956	hypothetical prote
2	436.5	43.3	170	H95088	conserved hypothet
3	331.5	32.9	251	B86647	hypothetical prote
4	241.5	23.9	127	F89858	conserved hypothet
5	239.5	23.7	91	A81375	B. subtilis YutD p
6	239.5	23.7	91	A81744	B. subtilis YutD p
7	234.5	23.2	93	A84079	hypothetical prote
8	232.5	23.0	102	F70023	hypothetical prote
9	109.5	10.9	1264	S84146	probable membrane
10	105	10.4	553	T27245	hypothetical prote
11	105	10.4	659	T27246	hypothetical prote
12	101	10.0	508	E71620	hypothetical prote
13	100.5	10.0	568	H88904	protein Y5G1LC.9
14	100	9.9	269	B47112	growth response pr
15	99.5	9.9	272	S59042	splicing factor SR
16	99	9.8	484	A40988	54K arginine-rich
17	99	9.8	494	A48133	pre-mRNA splicing
18	99	9.8	535	T37189	hypothetical prote
19	99	9.8	754	JC5314	CDC28/cdc2-like ki
20	95	9.4	350	A40459	nuclear phosphop
21	95	9.4	712	T25438	hypothetical prote
22	94.5	9.4	755	S58718	probable nuclear p
23	94.5	9.4	1650	T18444	hypothetical prote
24	93	9.2	1294	T18473	hypothetical prote
25	93	9.2	1507	B47328	natural killer cel
26	92	9.1	344	S59043	spilling factor SR
27	91.5	9.1	451	E70013	hypothetical prote
28	91.5	9.1	1403	A47328	natural killer cel
29	90.5	9.0	707	S57157	hypothetical prote

30	90.5	9.0	1203	2	S27545	pullulanase - Ther
31	90	8.9	462	2	C84473	probable protein k
32	90	8.9	470	2	T34402	hypothetical prote
33	90	8.9	2573	2	D71614	hypothetical prote
34	89.5	8.9	374	2	A37282	52K active chroma
35	89.5	8.9	1827	2	T16270	hypothetical prote
36	89	8.8	443	2	C88427	protein R07E5.6 [i
37	89	8.8	592	2	S43597	coiled-coil protei
38	89	8.8	686	1	A44842	CGMP-gated ion cha
39	89	8.8	846	2	JC7720	acetyltransferase
40	89	8.8	1788	2	T31095	vitellogenin precu
41	89	8.8	1997	2	F71607	DNA helicase II BR
42	88.5	8.8	382	1	VHIHPC	nucleocapsid prote
43	88.5	8.8	690	2	A42161	CGMP-gated cation
44	88	8.7	490	2	T43745	clt4 protein - fls
45	87.5	8.7	299	2	T25596	hypothetical prote

ALIGNMENTS

RESULT 1

C97956

hypothetical protein spr0675 [imported] - Streptococcus pneumoniae (strain R6)

C:Species: Streptococcus pneumoniae

C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001

C:Accession: C97956

R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S. Y, P.; Sun, P.M.; Winkler, M.E.

J. Bacteriol. 183, 5709-5717, 2001

A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S. A.

A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.

A:Reference number: A97872; MUID:21429245; PMID:11544234

A:Accession: C97956

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-176 <KUR>

A:Cross-references: GB:AE007317; PIDN:AAK99479.1; PID:g15458263; GSPDB:GN00174

C:Genetics:

A:Gene: spr0675

Query Match 44.8%; Score 452.5; DB 2; Length 176;

Best Local Similarity 47.6%; Pred. No. 2.7e-29;

Matches 91; Conservative 35; Mismatches 50; Indels 15; Gaps 4;

Qy 1 MRKEVTPMLNKNYPGPQFIHFENIVKSDDETFOLVINEKSAFDVTVFGQRFSEILUKY 60

Db 1 MRKEAPELYNKNYPGPPEFHLHGDKVETEGIAFSLVENIKDAFDVTFNQRFSEVLTKF 60

Qy 61 DFIVGDNQNEOLRGRFYKDASTIRKNSRISLEDIYKEYCNFGCAYFVLENPNDIKF 120

Db 61 DFIVGDNQNEQLRGRFYKDDRTKLEKISRLQDYLLEYCSYGCAFYVLENAEPRKRAF 120

Qy 121 DDEREHKRRKRSKSSQSSKTSOTRNRSQSNANAHFTSKRKDKTKRRQRHIEEQDKENT 180

Db 121 D--KFWKTEETPSKGGKKPTQTKR-KSNAD-----KNRRRQKDQHSQKEDKE-- 167

Qy 181 SAKQELLFVRK 191

Db 168 --QRFVIRQK 176

RESULT 2

H95088

conserved hypothetical protein sp0767 [imported] - Streptococcus pneumoniae (strain 1

C:Species: Streptococcus pneumoniae

C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001

C:Accession: H95088

R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzappel nson, T.; Hickley, E.K.; Holt, I.E.

Science 293, 498-506, 2001

THIS PAGE BLANK (USPTO)

```
FT DOMAIN 989 999 PRO-RICH.
FT DOMAIN 1351 1357 PRO-RICH.
FT DOMAIN 2094 2098 CORNR BOX OF ID1.
FT DOMAIN 2296 2300 CORNR BOX OF ID2.
FT DOMAIN 494 507 POLY-GLN.
FT DOMAIN 1615 1619 POLY-ALA.
FT DOMAIN 2434 2437 POLY-PRO.
FT VARSPLIC 36 254 MISSING (IN ISOFORM BETA).
FT CONFLICT 176 176 M -> RL (IN REF. 2).
FT CONFLICT 396 402 PPMLYDA -> RHVVR (IN REF. 2).
FT CONFLICT 555 555 D -> H (IN REF. 1; AAD20944).
FT CONFLICT 756 756 T -> M (IN REF. 1; AAD20944).
FT CONFLICT 785 785 V -> A (IN REF. 2).
FT CONFLICT 806 846 HHLPHRLMTRMNMKRLQLQPRMPRSRLPRSRSMWE
      -> PSPAAPATVDKDEQEAAPAPOTEDAKEQKSEAEI
      DVG (IN REF. 2).
      E -> K (IN REF. 1; AAD20945).
      E -> K (IN REF. 1; AAD20945).
      E -> K (IN REF. 1; AAD20945).
      S -> F (IN REF. 1; AAD20944).
      I -> IQ (IN REF. 1; AAD20944).
      PKLTEPRRWSGLPFPI -> QSYRLSPHAGHRLPSH
      (IN REF. 2).
      PHAADPSA -> TRADPL (IN REF. 2).
      MISSING (IN REF. 2).
      MISSING (IN REF. 2).
      G -> E (IN REF. 2).
      GSATSGSITKGLPSTRAADGPPSYRGSIHTG -> APPPVEA
      SPRASQYPCRRPQLQRLYHPR (IN REF. 2).
      A -> S (IN REF. 2).
      MISSING (IN REF. 2).
      A -> P (IN REF. 2).
      A -> G (IN REF. 2).
      G -> A (IN REF. 2).
      N -> S (IN REF. 2).
      A -> G (IN REF. 2).
      TA -> AV (IN REF. 2).
      LE -> SK (IN REF. 2).
      T -> A (IN REF. 2).
      2472 AA; 270856 MW; 2A58F4DF7B79285B CRC64;

Query Match 8.4%; Score 84.5; DB 1; Length 2472;
Best Local Similarity 19.9%; Pred. No. 50;
Matches 28; Conservative 31; Mismatches 47; Indels 35; Gaps 6;

Qy 72 LRLRGFYKDASTIRKNSRSL-----EDYKEYCNFG-----CA-- 106
Db 410 INMGLMDPMKYKDRQVTNNMWSQERDTERKXFMQHPKNGFLIASPLERKTVACVLY 469
Qy 107 -YFVLENPNDIKFDDERPHKRKRSKSSQSSKSSQTSOTNNRSQSNANAHTSKRKKDTR 165
Db 470 YLTKKNENYKSLV---RRSYRR---RGKSOQQQQQQQQQQQQOMAR---SSQEEKEKE 520
Qy 166 QOERHIKEEQDKEMTSAKOHL 186
Db 521 KEKEADKEEKEQDAENEKEEL 541
```

Search completed: February 26, 2003, 09:53:42
Job time : 19 secs


```

AC P07789; O05222;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Spore coat protein B.
GN COB.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98015417; PubMed=9353933;
RA Prescan E., Moszer I., Boursier L., Cruz Ramos H.C., De La Fuente V.,
RA Hullo M.-F., Lelong C., Schleich S., Sekowska A., Song B.H.,
RA Villani G., Kunst F., Danchin A., Glaser P.;
RT "The Bacillus subtilis genome from gerBC (311 degrees) to lncR (334
RT degrees).";
RL Microbiology 143:3313-3328(1997).
RN [2]
RP SEQUENCE OF 1-47 FROM N.A.
RC STRAIN=PV17;
RX MEDLINE=68011308; PubMed=2821284;
RA Donovan W., Zheng L., Sandman K., Losick R.;
RT "Genes encoding spore coat polypeptides from Bacillus subtilis.";
RN [3]
RP SEQUENCE OF 1-23 FROM N.A.
RX MEDLINE=68011308; PubMed=2821284;
RA Donovan W., Zheng L., Sandman K., Losick R.;
RT "Genes encoding spore coat polypeptides from Bacillus subtilis.";
RN [3]
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z93767; CAB07789.1; -
DR EMBL; X98342; -; NOT_ANNOTATED_CDS.
DR EMBL; X05679; CAA29166.1; -
DR EMBL; Z99122; CAB15622.1; -
DR PIR; B27393; B27393.
DR Subtilist; BG10491; cotB.
DR Sporulation; Complete proteome.
FT CONFLICT 44 47 AVKS -> GPAA (IN REF. 2).
SQ SEQUENCE 380 AA; 42971 MW; A42451945976CC79 CRC64;

Query Match 8.5%; Score 86; DB 1; Length 380;
Best Local Similarity 25.1%; Pred. No. 4.9;
Matches 49; Conservative 33; Mismatches 67; Indels 46; Gaps 11;

Qy 3 KEVTEMLNKNYKPGQPIHFNFI-VKSDDIIEFOLVINEKSAFDVTVFGQRFSEILLKYD 61
Db 173 KSLTKHWVSNR-GPEAI--EGILVDNADGHYTVKQ-----EVLRIYP 215

Qy 62 FIV-----GDWNEQLRLRGYKDASTIRKNSIRSLRLEYIKYCNFCGCAVFLNP 113
Dy 216 FHIKISILGPKGYSKKEDQKNEQEDNNDKDSNFTSS-----KSYSS-----SKS 262

Qy 114 NPRIKFDDEPHRR-RKRSKSSKSSKSSOTNRNRS---QSNANAHFTSKKRKDKTKRQER 169
Dy 263 SKRLKSSDDOSSKGRSRSKSS-SKSKSKRLKSSDYQSSKSGR--SSRSKSSKSSKSKR 319

Qy 170 HIKEEQDKEMTSKQ 184
Dy 320 SLKSSDYQSSKSKR 334

RESULT 36
NSRI_YEAST
ID NSRI_YEAST STANDARD; PRT; 414 AA.

```

```

AC P27476;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nuclear localization sequence binding protein (P67).
GN NSRI OR YGR159C OR G7001.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=91177946; PubMed=1706724;
RA Lee W.-C., Xue Z., Metlese T.;
RT "The NSRI gene encodes a protein that specifically binds nuclear
RT localization sequences and has two RNA recognition motifs.";
RL J. Cell Biol. 113:1-12(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=92355583; PubMed=1644811;
RA Kondo K., Inouye M.;
RT "Yeast NSRI protein that has structural similarity to mammalian
RT nucleolin is involved in pre-rRNA processing.";
RL J. Biol. Chem. 267:16252-16258(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=96158062; PubMed=8585325;
RA Skala J., Nawrocki A., Goffeau A.;
RT "The sequence of a 27 kb segment on the right arm of chromosome VII
RT from Saccharomyces cerevisiae reveals MOLL, NAT2, RPL30B, RSR1, CYS4,
RT PML/CHO2, NSRI genes and ten new open reading frames.";
RL Yeast 11:121-1427(1995).
RN [4]
RP SEQUENCE OF 188-414 FROM N.A.
RA Rieger M., Mueller-Auer S., Brueckner M., Schaefer M.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [5]
RP DNA-BINDING.
RX MEDLINE=95098604; PubMed=7800479;
RA Lin J.-J., Zakian V.A.;
RT "Isolation and characterization of two Saccharomyces cerevisiae genes
RT that encode proteins that bind to (TGI-3)n single strand telomeric
RT DNA in vitro.";
RL Nucleic Acids Res. 22:4906-4913(1994).
CC -!- FUNCTION: INVOLVED IN PRE-RRNA PROCESSING. SPECIFICALLY BINDS
CC NUCLEA? LOCALIZATION SEQUENCES. CANDIDATE FOR A RECEPTOR AT THE
CC NUCLEUS THAT MAY BE INVOLVED IN BOTH RNA AND PROTEIN TRANSPORT.
CC BINDS TELOMERIC SEQUENCES OF THE TYPE (TG(1-3)N IN VITRO.
CC -!- SUBCELLULAR LOCATION: NUCLEAR; POSSIBLY AT THE NUCLEOLUS.
CC -!- INDUCTION: IN RESPONSE TO LOW TEMPERATURE (BY COLD-SHOCK).
CC -!- SIMILARITY: BELONGS TO THE GAR FAMILY.
CC -!- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X57185; CAA40472.1; -
DR EMBL; X85607; CAA59817.1; -
DR EMBL; Z72544; CAA97173.1; -
DR EMBL; Z72546; CAA97180.1; -
DR PIR; A392C5; A39205.
DR SGD; S0003391; NSRI.
DR InterPro; IPR000504; RNA_rec_mot.
DR SMART; PF00076; rrm; 2.
DR SMART; SM0360; RRM; 2.

```

```

QY 173 EQDKEMTSKAK 184
ID YF13_YEAST STANDARD; PRT: 787 AA.
Db 689 EERGRKSKSRK 700

RESULT 33
YF13_YEAST
ID YF13_YEAST STANDARD; PRT: 787 AA.
AC P43596;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Hypothetical 90.9 kba protein in GCN20-CMK1 intergenic region.
GN YF013W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RX MEDLINE=95400292; PubMed=7670463;
RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
RA Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
RA Yamazaki M., Tashiro H., Eki T.;
RT "Analysis of the nucleotide sequence of chromosome VI from
RT Saccharomyces cerevisiae."
RL Nat. Genet. 10:261-268(1995).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D50617; BAA09252.1;
DR SGD; S0001909; IOC3.
KW Hypothetical protein.
SQ SEQUENCE 787 AA; 90896 MW; 1919A8A1F58B5340 CRC64;

Query Match 8.6%; Score 86.5; DB 1; Length 787;
Best Local Similarity 24.1%; Pred. No. 9.8;
Matches 27; Conservative 22; Mismatches 42; Indels 21; Gaps 4;

QY 69 NEQLRLRGYKDA--STIRKNSRISRLDYIKCYCNFGCAYFVLEPNPRDIKFDDEPH 126
Db 682 NPSINTNPLPKAKYNTARKKIQI--LKFLSDY-----YFILRFQFMKVFQADMKPG 733
QY 127 KRRKRSKSQSSKSTQTRNNRSQSNANAHFTSKRRKDTKRRQERHIKEEQDKE 178
Db 734 KROLRIQRT-----VNYTEYDSEYVDDEEDDEADIYDDNDND 774

RESULT 34
DDX8_HUMAN
ID DDX8_HUMAN STANDARD; PRT: 1220 AA.
AC Q14562;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ATP-dependent helicase DDX8 (RNA helicase HRH1) (DEAH-box protein 8).
GN DDX8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95021291; PubMed=7935475;
RA Ono Y., Ono M., Shimura Y.;

"Identification of a putative RNA helicase (HRH1), a human homolog of
yeast Prp22."
Mol. Cell. Biol. 14:7611-7620(1994).
[2]
RP FUNCTION.
RX MEDLINE=96196415; PubMed=8608946;
RA Ohno M., Shimura Y.;
RT "A human RNA helicase-like protein, HRH1, facilitates nuclear export
RT of spliced mRNA by releasing the RNA from the spliceosome."
RL Genes Dev. 10:997-1007(1996).
CC -!- FUNCTION: FACILITATES NUCLEAR EXPORT OF SPLICED MRNA BY RELEASING
CC THE RNA FROM THE SPLICEOSOME.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- DOMAIN: THE RS DOMAIN CONFERS A NUCLEAR LOCALIZATION SIGNAL, AND
CC APPEARS TO FACILITATE THE INTERACTION WITH THE SPLICEOSOME.
CC -!- SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY. DEAH
CC SUBFAMILY DDX8/PRP22 ORTHOLOG.
CC -!- SIMILARITY: CONTAINS 1 SI MOTIF DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D50487; BAA09078.1;
DR HSP; P05055; LSRO.
DR Genew; HGNC:2749; DDX8.
DR MIM; 600396;
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002464; DEAH_box.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR003029; SI.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF00575; SI; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELICC; 1.
DR SMART; SM00316; SI; 1.
DR PROSITE; PS00690; DEAH-ATP_HELICASE; 1.
DR PROSITE; PS0126; SI; 1.
KW mRNA processing; mRNA splicing; Helicase; ATP-binding;
KW Nuclear protein.
FT DOMAIN 172 175 POLY-LYS.
FT DOMAIN 176 228 ARG/SER-RICH (RS DOMAIN).
FT DOMAIN 265 336 SI MOTIF.
FT NP_BIND 588 595 ATP (POTENTIAL).
FT SITE 685 688 DEAH BOX.
FT MUTAGEN 594 594 K->E: IN GET; INHIBITION OF PRE-MRNA
FT SPlicing AND NUCLEAR EXPORT OF UNSPLICED
FT RNA.
FT MUTAGEN 717 717 S->L: IN LAT; INHIBITION OF PRE-MRNA
FT SPlicing AND NUCLEAR EXPORT OF UNSPLICED
FT RNA.
SQ SEQUENCE 1220 AA; 139314 MW; 17C1602A73A0EF24 CRC64;

Query Match 8.6%; Score 86.5; DB 1; Length 1220;
Best Local Similarity 35.3%; Pred. No. 16;
Matches 24; Conservative 18; Mismatches 21; Indels 5; Gaps 3;

QY 116 RDIKFDDERPHKRR-KSRSKSQSSKSTQTRN---NRSQSNANAHFTSKRRKDTKRRQERHI 171
Db 188 RDRERNRDRDHKRRHSRSR-SRTREKNVKSRYSRSRSQSPPKDRKDRDKYGERNL 246
QY 172 KEEQDKEM 179
Db 247 DRWRDKHV 254

RESULT 35
COTB_BACSU
ID COTB_BACSU STANDARD; PRT: 380 AA.

```

```

ID CYLL_HUMAN STANDARD; PRT; 598 AA.
AC P35663;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DE Cyclicin I (Multiple-band polypeptide I) (Fragment).
GN CYLC1 OR CYL1 OR CYL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Br15:col N2;
RX MEDLINE=93359502; PubMed=8354692;
RA Hess H., Heid H., Franke W.W.;
RT "Molecular characterization of mammalian cyclicin, a basic protein of
the sperm head cytoskeleton."
RL J. Cell Biol. 122:1043-1052(1993).
CC -!- FUNCTION: POSSIBLE ARCHITECTURAL ROLE DURING SPERMATOGENESIS. MAY
BE INVOLVED IN SPERMATID DIFFERENTIATION.
CC -!- SUBCELLULAR LOCATION: CYLIX; SPERM HEAD CYTOSKELETAL STRUCTURE.
CC -!- TISSUE SPECIFICITY: TESTIS.
CC -!- DEVELOPMENTAL STAGE: SPECIFIC TO LATE SPERMATOGENESIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z22780; CAA80457.1; -.
DR PIR; S35920; S35920.
DR PIR; B40713; B40713.
DR Genew; HGNC:2582; CYLC1.
DR MIM; 603121; -.
KW Cytoskeleton; Structural protein; Repeat; Sperm; Spermatogenesis.
FT NON_TER 1 1
FT DOMAIN 479 500 8 APPROXIMATE TANDEM REPEATS.
FT REPEAT 225 252 1.
FT REPEAT 253 289 2.
FT REPEAT 290 326 3.
FT REPEAT 327 364 4.
FT REPEAT 365 400 5.
FT REPEAT 401 438 6.
FT REPEAT 439 478 7.
FT REPEAT 479 500 8.
FT DOMAIN 548 598 PRO-RICH.
FT SEQUENCE 598 AA; 68034 MW; 062BA2E2D2D2AB61F7 CRC64;

Query Match 8.6%; Score 86.5; DB 1; Length 598;
Best Local Similarity 21.6%; Pred. No. 7.3;
Matches 38; Conservative 38; Mismatches 71; Indels 29; Gaps 7;

QY 13 NKYPGPGFIHFENIVKSDIEFOLVINEKSAFDVTVFGQRFSEILLKYDF--IVGDWGN 70
Db 153 SKTNTFELTKNNPKKD-----LKRKTSND--PISEICSENLSNVDLMLVGQSDDE 204
QY 71 QLRLRGYKDASTIRKNSIRSLRDEYKEYCNF-----GCAYFVLENPNDRIKFD 122
Db 205 SINFDALRWYS--QNSK-----NYSKYTKYTKTKTKNAKSSDAESDSKDAKDS 257
QY 123 ERPHKRRKSKSSQSTNNRQSNANAHFTSKRKDKTKRQERHIEEQDKE 178
Db 258 KVKVKNVKKDDKKDKKDKTESTDAESG----SKDERKDKTKDKKKLKKDKKKD 309

RESULT 32
Y061_CAEEL
ID Y061_CAEEL
AC P34600; STANDARD; PRT; 724 AA.

DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothesized. 84.7 kDa protein ZK1098.1 in chromosome III.
GN ZK1098.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Br15:col N2;
RX MEDLINE=94130718; PubMed=7906398;
RA Willson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laisster N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkhen R.,
RA Sims M., Smaildon N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstein L., Wilkinson-Sproat J.,
RA Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans."
RL Nature 368:32-38(1994).
CC -!- SIMILARITY: CONTAINS 2 WW DOMAINS.
CC -!- SIMILARITY: TO S.POMBE SPAC4D7.13.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z22176; CAA80142.1; -.
DR PIR; S40923; S40923.
DR WormPep; ZM1098.1; CE03847.
DR InterPro; IPR002713; FF.
DR InterPro; IPR001202; WW_Rsp5_WWP.
DR Pfam; PF00397; WW; 2.
DR Pfam; PF01446; FF; 5.
DR SMART; SM00441; FF; 3.
DR SMART; SM00456; WW; 2.
DR PROSITE; P01159; WW_DOMAIN_1; 2.
DR PROSITE; P50020; WW_DOMAIN_2; 2.
KW Hypothetical protein; Repeat.
FT DOMAIN 78 111 WW 1.
FT DOMAIN 123 156 WW 2.
FT SEQUENCE 724 AA; 84664 MW; BCC7AACAD5AA4C20 CRC64;

Query Match 8.6%; Score 86.5; DB 1; Length 724;
Best Local Similarity 17.5%; Pred. No. 9;
Matches 44; Conservative 45; Mismatches 84; Indels 79; Gaps 8;

QY 2 RKEVTPMLNKNYKPGQFIHFENIVKSDIEFOLVINEKSAFDVTVF----- 49
Db 459 RGLTMSLWTSLFP-----IISTDTRFELMLFQPGSSPLDLFKFFVEDLKEQYT 508
QY 50 --GQRTSEILLK-----YDFIVG-----DWGNQLRLRGFYKDA----- 81
Db 509 EDRRLIKELITEKGCQVIATTEYREFSDVWVSHKGGKVDHGNMKLCYNLSLIEAKSKAK 568
QY 82 ----STIRKNSRI--SRLEDYIKEY-----CNFGCAYFVLENPNDRIKFD 122
Db 569 DEEKESLRKRRLSEFEFNLLKEHNVDKDSWTVIKPIEKDKAYLAMENDERETAFNH 626
QY 123 ERPHK-----RRKRSKSSQSTNNRQSNANAHFTSKRKDKTKRQERHIEQ 172
Db 629 YKNGTSGTTAGSETLEKRRKKKKDKKKRSDNNSESEGEIREKREKKKKKKHSDRMD 688
```



```
DR InterPro: IPR001680; WD40.
DR Pfam: PF00440; WD40; 7.
DR PRINTS: PF02816; MHCK_EF2_Kinase; 1.
DR PRODOM: GPRO0320; GPROTEINBRPT.
DR PRODOM: PD000018; WD40; 6.
DR SMART: SM00320; WD40; 7.
DR PROSITE: PS00678; WD_REPEATS_1; 5.
DR PROSITE: PS00082; WD_REPEATS_2; 6.
DR PROSITE: PS0294; WD_REPEATS_REGION; 1.
KW transferase; Serine/threonine-protein kinase; ATP-binding; Repeat;
KW WD repeat.
FT NP_BIND 298 303 ATP (POTENTIAL).
FT DOMAIN 355 381 POLY-ASN.
FT REPEAT 458 486 WD 1.
FT REPEAT 500 528 WD 2.
FT REPEAT 540 568 WD 3.
FT REPEAT 580 608 WD 4.
FT REPEAT 620 648 WD 5.
FT REPEAT 660 688 WD 6.
FT REPEAT 700 730 WD 7.
SQ SEQUENCE 732 AA; 83167 MW; A7233C4BD56D4088 CRC64;

Query Match 8.6%; Score 87; DB 1; Length 732;
Best Local Similarity 20.9%; Pred. No. 8.3;
Matches 45; Conservative 29; Mismatches 69; Indels 72; Gaps 11;

QY 12 YNKYPG-----PQ-FTHFENIVKSDIEFOLVINEKSAFDTVFGQRFSEILL-KY 60
DB 242 YNNSGFVSNDRNTPQSFHSF-----TYEHSNHQLLI-----IDIQGVGDHYTDPQIHTY 292
QY 61 D---FIVGDWNEQLRLRGFYKDASTIRKNSRISRLDYIKCYNFCGCAVEFLNPNPR- 116
DB 293 DGVGFGIGNLQG-----KGFEKFLDTHK-----CNAICQYLNLOSINPKS 332
QY 117 -----DIKFDDEPHKRKRKSKSQSSKSSQTRNNRSQSNANAHETSKK-----R 160
DB 333 EKSDCGTVPRDLFPDTSERDNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 392
QY 161 KDTFRERHRIKEQ-----DKEWTSK 183
DB 393 SGSKERNDROSPSRQLFVSDNGTNTLNKERSKSK 427

RESULT 27
MN4_YEAST STANDARD; PRT; 1178 AA.
AC P36044; P36043; P89095;
DT 01-JUN-1994 (Rel. 29, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE MN4 protein.
GN MN4 OR YKL200C/YKL201C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c;
RX MEDLINE=971175967; PubMed=9023541;
RA Odani T., Shimizu Y.-I., Tanaka A., Jigami Y.;
RT "Cloning and analysis of the MN4 gene required for phosphorylation
of N-linked oligosaccharides in Saccharomyces cerevisiae."
RL Glycobiology 6:805-810(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Maia e Silva A., Bossier P., Vilela C., Fernandes L., Soares H.,
RA Guerreiro P., Rodrigues-Pousada C.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAY FUNCTION AS A POSITIVE REGULATOR FOR
CC MANNOSYLPHOSPHATE TRANSFERASE. IS REQUIRED TO MEDIATE
CC MANNOSYLPHOSPHATE TRANSFER IN BOTH THE CORE AND OUTER CHAIN
CC PORTIONS OF N-LINKED OLIGOSACCHARIDES.
```

CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC -1- SIMILARITY: TO YEAST YJR061W.
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO
CC FRAMESHIFTS, ONE OF WHICH PRODUCES TWO SEPARATE ORFS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isp-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: D83006; BAAL1676.1; -;
DR EMBL: Z28201; CAA82046.1; -;
DR EMBL: Z28200; CAA82044.1; -;
DR PIR: S38037; S38037.
DR PIR: S38038; S38038.
DR SGD: S0001683; MN4.
KW Transmembrane; Signal-anchor; Repeat.
FT DOMAIN 1 27 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 28 48 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
(POTENTIAL).
FT DOMAIN 49 1178 LUMENAL (POTENTIAL).
FT DOMAIN 1032 1174 ARG/GLU/LYS-RICH (HIGHLY CHARGED).
FT DOMAIN 1042 1174 17 X 8 AA TANDEM REPEATS OF K-K-K-E-E-
E-E.
FT REPEAT 1042 1049 1.
FT REPEAT 1050 1057 2.
FT REPEAT 1058 1065 3.
FT REPEAT 1066 1073 4.
FT REPEAT 1074 1081 5.
FT REPEAT 1082 1089 6.
FT REPEAT 1090 1097 7 (APPROXIMATE).
FT REPEAT 1098 1105 8.
FT REPEAT 1106 1113 9 (APPROXIMATE).
FT REPEAT 1114 1121 10 (APPROXIMATE).
FT REPEAT 1122 1129 11 (APPROXIMATE).
FT REPEAT 1130 1137 12.
FT REPEAT 1138 1144 13 (APPROXIMATE).
FT REPEAT 1145 1152 14 (APPROXIMATE).
FT REPEAT 1153 1160 15 (APPROXIMATE).
FT REPEAT 1161 1168 16 (APPROXIMATE).
FT REPEAT 1169 1174 17 (INCOMPLETE).
FT DOMAIN 37 40 POLY-ILE.
FT DOMAIN 1042 1045 POLY-LYS.
FT DOMAIN 1046 1049 POLY-GLU.
FT DOMAIN 1050 1053 POLY-LYS.
FT DOMAIN 1054 1057 POLY-GLU.
FT DOMAIN 1058 1061 POLY-LYS.
FT DOMAIN 1062 1065 POLY-GLU.
FT DOMAIN 1066 1069 POLY-LYS.
FT DOMAIN 1070 1073 POLY-GLU.
FT DOMAIN 1074 1077 POLY-LYS.
FT DOMAIN 1078 1081 POLY-GLU.
FT DOMAIN 1082 1085 POLY-LYS.
FT DOMAIN 1086 1089 POLY-GLU.
FT DOMAIN 1094 1097 POLY-GLU.
FT DOMAIN 1098 1101 POLY-LYS.
FT DOMAIN 1102 1105 POLY-GLU.
FT DOMAIN 1134 1137 POLY-GLU.
FT DOMAIN 1157 1160 POLY-GLU.
FT DOMAIN 1165 1168 POLY-GLU.
SQ SEQUENCE 1178 AA; 139380 MW; BC05DAE0AEFCB282 CRC64;
Query Match 8.6%; Score 87; DB 1; Length 1178;
Best Local Similarity 18.7%; Pred. No. 14;
Matches 34; Conservative 34; Mismatches 76; Indels 38; Gaps 5;
QY 19 QFIHFENIVKSDIEFOLVINEKSAFDTVFGQRFSEILLKYDFIVGDWNEQLRLRGFY 78
DB 924 KIVHEEVGVGIIDLSLSKYNFSLFNIT-----FDETGTTLDDNTED-----Y 967

DR Pfam; PF00642; zf-COCH; 2.
DR SMART; SM00360; RRM; 1.
DR SMART; SM00366; Znf_C3H1; 2.
DR PROSITE; PS0102; RRM; 1.
DR PROSITE; PS00030; RRM_RNP1; FALSE_NEG.
KW Nuclear protein; RNA-binding; Ribonucleoprotein; Zinc-finger; Repeat.
FT DOMAIN 189 295 RNA-BINDING (RRM).
FT DOMAIN 85 88 POLY-LYS.
FT DOMAIN 116 119 POLY-GLU.
FT CONFLICT 351 351 N -> S (IN REF. 2).
SQ SEQUENCE 428 AA; 51364 MW; 4542EE234B0759E6 CRC64;

Query Match 8.7%; Score 87.5; DB 1; Length 428;
Best Local Similarity 27.0%; Pred. No. 4.2;
Matches 31; Conservative 9; Mismatches 44; Indels 31; Gaps 5;

QY 99 EYCNFGCAVYVLENPN-----PRDI-----KFDDEPHKRR-----KRSK 134
DB 315 KHCNF---LHVFPNPNNEPREANRDYMSPPAWTGGSGKNSDRERKDHHEEYYSKRSY 371
QY 135 SOSKSQTRNNSQSNANAFTSKKRKDKRKQERHI-----KEEQDKEMTSAKOH 185
DB 372 HSGSYHSSARNRESKSPHRKKSHKQTKSHERRSSRRGREDSPPGPQSOSH 426

RESULT 25
EX5B_BORBU
ID EX5B_BORBU STANDARD; PRT; 1169 AA.
AC 051578;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Exodeoxyribonuclease V beta chain (EC 3.1.11.5).
GN RECB OR BB0633.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra B., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA van Vugt R., Palmer N., Adams M.D., Cocayne J.D., Weidman J.,
RA Uitterback T., Wathey L., McDonald L., Artlich P., Bowman C.,
RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi".
RL Nature 390:580-586(1997).
CC -!- FUNCTION: REQUIRED FOR EFFICIENT DNA REPAIR; IT CATALYZES THE UNWINDING OF DOUBLE-STRANDED DNA AND THE CLEAVAGE OF SINGLE-STRANDED DNA AND IT STIMULATES LOCAL GENETIC RECOMBINATION.
CC ALL OF THESE ACTIVITIES REQUIRE CONCOMITANT HYDROLYSIS OF ATP (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: Exonucleolytic cleavage (in the presence of ATP) in either 5'- to 3'- or 3'- to 5'-direction to yield 5'-phosphooligonucleotides.
CC -!- SUBUNIT: CONSIST OF THREE SUBUNITS; RECB, RECC AND RECD (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE HELICASE FAMILY. UVRD SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC EMBL; AE001164; AAC66981.1; -

DR HSSP; P56255; LPJR.
DR TIGR; BB0633;
DR InterPro; IPR004586; RecB.
DR InterPro; JPR000212; UvrD-helicase.
DR Pfam; PF0080; UvrD-helicase; 1.
DR TIGRFAMS; TIGR00609; recB; 1.
KW Hydrolyase; Nuclease; Exonuclease; Endonuclease; Helicase; ATP-binding; DNA repair; Complete proteome.
FT NP_BIND 18 25 ATP (POTENTIAL).
SQ SEQUENCE 1169 AA; 137828 MW; B61D63C1C959B91F CRC64;

Query Match 8.7%; Score 87.5; DB 1; Length 1169;
Best Local Similarity 23.1%; Pred. No. 13;
Matches 48; Conservative 32; Mismatches 77; Indels 51; Gaps 9;

QY 8 EMLNTYIKY-PGPQF-----IHFENIVKSDIETEFOLVINEKSAFDVTFV---GQRFSEIL 57
DB 109 ETENYSKYKPKKEFSKEIDEIVDFLRKSDSLIQALDIKD---YELKVFKSDARKTEIV 165
QY 58 LKYDF-----VGDAGNEQ-----LRLRGFYKDASTI-----RK 86
DB 166 LAIKKAYERTDQELGDMWKQTAFENILLKKEELIKDINKLIEDLDMKTDKDEILSFYNK 225
QY 87 NSRISKLE-DYIKCYCNFGCAVYVLENPNPRDIKEDDERPHKRRKRSKSSKSQTSQTRNN 145
DB 226 HIQTGKLEIEYSKENDIFKIAETLLKN-----KFFSTLIEKETKKNKSLSPKELKIKND 279
QY 146 RSQSNINAHFTSKKRRKDKRKQERHIKE 173
DB 280 LITGLINIKHEKYSKEDNRNKNRNLKQ 307

RESULT 26
KMH_B_DICDI
ID KMH_B_DICDI, STANDARD; PRT; 732 AA.
AC P90648;
DT 15-JUL-1993 (Rel. 36, Created)
DT 15-JUL-1993 (Rel. 36, Last sequence update)
DT 15-JUL-1993 (Rel. 36, Last annotation update)
DE Myosin heavy chain kinase B (EC 2.7.1.129) (MHCK B).
GN MHKB OR MH2KB.
OS Dictyostelium discoideum (slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX3;
RX MEDLINE=97277316; PubMed=9115238;
RT "Identification of a protein kinase from Dictyostelium with homology to the novel catalytic domain of myosin heavy chain kinase A.";
RL J. Biol. Chem. 272:11812-11815(1997).
CC -!- FUNCTION: PHOSPHORYLATES THREONINE IN THE C-TERMINAL TAIL REGION OF MYOSIN II HEAVY CHAIN. THIS PHOSPHORYLATION IS CRITICAL IN REGULATING THE ASSEMBLY AND DISASSEMBLY OF MYOSIN II FILAMENT.
CC -!- CATALYTIC ACTIVITY: ATP + [myosin heavy-chain] = ADP + [myosin heavy-chain] phosphate.
CC -!- DOMAIN: CONSISTS OF AN N-TERMINAL DOMAIN, A CENTRAL NONREPETITIVE CATALYTIC DOMAIN, AND A C-TERMINAL DOMAIN WITH SEVEN WD REPEATS.
CC -!- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
CC -!- SIMILARITY: BELONGS TO THE MHCK / EF-2 PROTEIN KINASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC EMBL; U90946; AAB50136.1; -
CC Dictyob; ID01087; mhkb.
DR InterPro; IPR004166; MHCK_EF2_kinase.

CC COUPLED CASCADE USING CGMP AS SECOND MESSENGER. THIS PROTEIN CAN
 CC BE ACTIVATED BY CYCLIC GMP WHICH LEADS TO A OPENING OF THE CATION
 CC CHANNEL AND THEREBY CAUSING A DEPOLARIZATION OF ROD
 CC PHOTORECEPTORS.
 CC
 CC -1- SUBUNIT: HOMOTETRAMER OR HIGHER OLIGOMER. FORMS HETEROOLIGOMERIC
 CC COMPLEX WITH CNGA.
 CC
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC
 CC -1- TISSUE SPECIFICITY: ROD CELLS IN THE RETINA.
 CC
 CC -1- DISEASE: DEFECTS IN CNGA1 ARE A CAUSE OF AUTOSOMAL RECESSIVE
 CC RETINITIS PIGMENTOSA (ARRP). A DISEASE THAT LEADS TO DEGENERATION
 CC OF RETINAL PHOTORECEPTOR CELLS.
 CC
 CC -1- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE-GATED CATION CHANNEL
 CC FAMILY.
 CC
 CC -1- DATABASE: NAME=Mutations of the CNGA1 gene;
 CC NOTE=Retina International's Scientific Newsletter;
 CC WWW="http://www.retina-international.com/sci-news/cngalmut.htm".
 CC
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M84741; AA52010.1; ALT_INIT.
 CC EMBL; S42457; AAB22778.1; -.
 CC EMBL; S76062; AAD14206.1; -.
 CC PIR; A42161; A42161.
 CC Genew; HGNC:2148; CNGA1.
 CC MTM; 123825; -.
 CC InterPro; IPR000636; M+channel_nlg.
 CC InterPro; IPR000595; cNMP_binding.
 CC Pfam; PF00027; cNMP_binding; 1.
 CC Pfam; PF00520; ion_trans; 1.
 CC SMART; SM00100; cNMP; 1.
 CC PROSITE; PS00888; CNMP_BINDING_1; 1.
 CC PROSITE; PS00889; CNMP_BINDING_2; 1.
 CC PROSITE; PS00042; CNMP_BINDING_3; 1.
 CC Ionic channel; Ion transport; cAMP-binding; Transmembrane;
 CC Multigene family; Vision; Disease mutation; Polymorphism;
 CC Retinitis pigmentosa.
 CC
 CC FT DOMAIN 1 160 CYTOPLASMIC (POTENTIAL).
 CC FT TRANSMEM 161 181 H1 (POTENTIAL).
 CC FT DOMAIN 182 194 EXTRACELLULAR (POTENTIAL).
 CC FT TRANSMEM 195 213 H2 (POTENTIAL).
 CC FT DOMAIN 214 237 CYTOPLASMIC (POTENTIAL).
 CC FT TRANSMEM 238 257 H3 (POTENTIAL).
 CC FT DOMAIN 258 295 EXTRACELLULAR (POTENTIAL).
 CC FT TRANSMEM 296 318 H4 (POTENTIAL).
 CC FT DOMAIN 319 370 CYTOPLASMIC (POTENTIAL).
 CC FT TRANSMEM 371 390 H5 (POTENTIAL).
 CC FT DOMAIN 391 474 EXTRACELLULAR (POTENTIAL).
 CC FT TRANSMEM 475 495 H6 (POTENTIAL).
 CC FT DOMAIN 496 685 CYTOPLASMIC (POTENTIAL).
 CC FT NP_BIND 483 605 CGMP (POTENTIAL).
 CC FT BINDING 542 542 CGMP (POTENTIAL).
 CC FT BINDING 557 557 CGMP (POTENTIAL).
 CC FT CARBOHYD 421 421 N-LINKED (GLCNAC...) (POTENTIAL).
 CC FT VARIANT 28 28 R -> Q.
 CC FT VARIANT 114 114 /FTId=VAR_009295.
 CC FT VARIANT 316 316 /FTId=VAR_009296.
 CC FT VARIANT 46 46 S -> F (IN ARR).
 CC FT CONFLICT 85 85 S -> Y (IN REF. 1).
 CC FT CONFLICT 146 147 EE -> HH (IN REF. 1).
 CC FT CONFLICT 539 539 Y -> T (IN REF. 1).
 CC FT CONFLICT 677 678 GA -> WS (IN REF. 1).
 CC SEQUENCE 686 AA; 79126 MW; E520D0216FC97AF6 CRC64;

Query Match

8.7%; Score 88; DB 1; Length 686;

Best Local Similarity 27.0%; Pred. No. 6.4;
 Matches 31; Conservative 23; Mismatches 41; Indels 20; Gaps 6;
 QY 69 NEQLRLRGFKDASTIRKNSRISLEDYIKEYCNFGCAYFVLEPNPRDIKFDDEPHKR 128
 Db 53 NENPHARGSFYSYKS-LRKGGPSQR-----EQYLPGATALFNVNSSNK-----DOEPEEK 101
 QY 129 RKRSKSSQSKSQTNRNRSQSNANAHFTSKRKDKTKRQE----RHIIKEEQDKEM 179
 Db 102 KK-KKKEKKSQSDKN----ENKNDPEKSKKKKKKKKEKSKSKKKEKKEKREK 151
 RESULT 24
 U2R1_MOUSE
 ID U2R1_MOUSE STANDARD; PRT; 428 AA.
 AC Q64707;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE U2 small nuclear ribonucleoprotein auxiliary factor 35 kDa subunit
 DE related-protein 1 (SP2).
 GN U2AF1-RS1 OR SP2 OR SP2-7.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RX SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=94111973; PubMed=8284201;
 RA Hatada I., Sugama T., Mukai T.;
 RT "A new imprinted gene cloned by a methylation-sensitive genome
 RT scanning method.";
 RL Nucleic Acids Res. 21:5577-5582(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PWK; TISSUE=Liver;
 RX MEDLINE=94184361; PubMed=8136831;
 RA Hayashizaki Y., Shibata H., Hirotsune S., Sugino H., Okazaki Y.,
 RA Sasaki N., Hirose K., Imoto H., Okuizumi H., Muramatsu M.,
 RA Komatsubara H., Shiroishi T., Moriwaki K., Katsuki M., Hatano N.,
 RA Sasaki H., Ueda T., Mise N., Takagi N., Plass C., Chapman V.M.;
 RT "Identification of an imprinted U2af binding protein related sequence
 RT on mouse chromosome 11 using the RUGS method.";
 RL Nat. Genet. 6:33-40(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=95175367; PubMed=7870588;
 RA Hatada I., Kitagawa K., Yamaoka T., Wang X., Arai Y., Hashido K.,
 RA Ohishi S., Masuda J., Ogata J., Mukai T.;
 RT "Allele-specific methylation and expression of an imprinted U2af1-rs1
 RT (SP2) gene.";
 RL Nucleic Acids Res. 23:36-41(1995).
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
 CC -1- SIMILARITY: CONTAINS 2 C3H1-TYPE ZINC FINGERS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D17407; BAA04230.1; -.
 CC EMBL; S69507; AAB30301.1; -.
 CC EMBL; D26474; BAA05486.1; -.
 CC MGD; MGI:98885; U2af1-rs1.
 CC InterPro; IPR000504; RNA_rec_mot.
 CC InterPro; IPR000571; Znf_CCCH.
 CC Pfam; PF00076; rrm; 1.

DR InterPro: IPR000237; GRIP_domain.
 DR Pfam: PF01465; GRIP; 1.
 KW Golgi stack; Antigen: Coiled coil; Alternative splicing.
 FT DOMAIN 133 237
 FT DOMAIN 276 1011
 FT DOMAIN 1033 1214
 FT DOMAIN 1259 2152
 FT VARSPLIC 2154 2185
 FT AICTIRMSHLENLPNLSICEKCFLSI (IN ISOFORM 2).
 FT MISSING (IN ISOFORM 2).
 FT VARSPLIC 2186 2230
 FT VARSPLIC 2103 2109
 FT VARSPLIC 2222 2230
 FT VARSPLIC 2222 2230
 FT FTSPRSIGF -> SWLRSSS (IN ISOFORM 4).
 FT R -> K (IN REF. 3).
 FT CONFLICT 188 188
 FT CONFLICT 220 220
 FT CONFLICT 220 220
 FT CONFLICT 276 276
 FT CONFLICT 584 584
 FT CONFLICT 628 628
 FT CONFLICT 630 630
 FT CONFLICT 682 682
 FT CONFLICT 682 682
 FT K -> N (IN REF. 3).
 SQ SEQUENCE 2230 AA; 261139 MW; 3BB733DB1EA86134 CRC64;
 Query Match 9.0%; Score 90.5; DB 1; Length 2230;
 Best Local Similarity 20.1%; Pred. No. 15;
 Matches 38; Conservative 47; Mismatches 75; Indels 29; Gaps 8;
 Qy 12 YNKPQPGQFIHFENIVKSDDDIEFQVLVINEKSAFD-----VTVFQGRFSEILLYKDYFIVGDW 67
 Db 1445 WKKAQSRFTQHTQTVK--ELQIOLEKSKYAEKDEQINLLKELDOONKRFDCCLKGEM 1502
 Qy 68 GNEQLRLRGFYKDASTRKN--SRISRLIEDYI-KEYCNFGCAYFVLEPNP-PRDIKEDDE 123
 Db 1503 EDDSKMEKESNLETLEKLSQATRI MELEDHTOKTIEISLNEVLKYNQOKDIE----- 1558
 Qy 124 RPHKRRSRKSSQSKSOTRRNSQ-----SNANAHTSKRKDTRKRQERHI----- 171
 Db 1559 --HKQVLQKLFQELGEEKDNVRKAEKILITLENQVYSMAKAELETKKKELEHVNLSVK 1616
 Qy 172 -KEQDKEM 179
 Db 1617 SKEBELKAL 1625
 RESULT 20
 NCR2_HUMAN
 ID NCR2_HUMAN STANDARD; PRT; 2517 AA.
 AC Q9Y618; Q9Y500; Q13354; O00613; O15416;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Nuclear receptor co-repressor 2 (N-CoR2) (Silencing mediator of
 DE retinoic acid and thyroid hormone receptor) (SMRT) (SMRte) (Thyroid-,
 DE retinoic-acid-receptor-associated co-repressor) (T3 receptor-
 DE associating factor) (TRAC) (CTG26).
 GN NCR2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM SMRT).
 RC TISSUP=Pituitary;
 RX MEDLINE=99178941; PubMed=10077563;
 RA Ordenlich P., Downes M., Xie W., Genin A., Spinner N.B., Evans R.M.;
 RL "Unique forms of human and mouse nuclear receptor corepressor SMRT";
 RT Proc. Natl. Acad. Sci. U.S.A. 96:2639-2644(1999).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM SMRT).
 RC TISSUP=Cervical adenocarcinoma;
 RX MEDLINE=99199215; PubMed=10097088;
 RA Park E.J., Schroeen D.J., Yang M., Li H., Li L., Chen J.D.;
 RT "SMRte, a silencing mediator for retinoid and thyroid hormone
 RT receptors-extended isoform that is more related to the nuclear

receptor corepressor.";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:3519-3524(1999).
 RN [3]
 RP SEQUENCE OF 1023-2517 FROM N.A.
 RC TISSUP=Cervical adenocarcinoma;
 RX MEDLINE=96008552; PubMed=7566127;
 RA Chen J.D., Evans R.M.;
 RT "A transcriptional co-repressor that interacts with nuclear hormone
 RT receptors.";
 RL Nature 377:454-457(1995).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM TRAC-1).
 RC TISSUP=Fetal liver;
 RX MEDLINE=96408715; PubMed=8813722;
 RA Sande S., Privalsky M.L.;
 RT "Identification of TRACs (T3 receptor-associating cofactors), a family
 RT of cofactors that associate with, and modulate the activity of,
 RT nuclear hormone receptors.";
 RL Mol. Endocrinol. 10:813-825(1996).
 RN [5]
 RP SEQUENCE OF 428-613 FROM N.A.
 RC TISSUP=Brain cortex;
 RX MEDLINE=97369492; PubMed=9225980;
 RA Margolis R.L., Abraham M.R., Gatchell S.B., Li S.H., Kidwai A.S.,
 RA Breschel T.S., Stine O.C., Callahan C., McInnis M.G., Ross C.A.;
 RT "cDNAs with long CAG trinucleotide repeats from human brain.";
 RL Hum. Genet. 100:114-122(1997).
 CC -1- FUNCTION: MEDIATES THE TRANSCRIPTIONAL REPRESSION ACTIVITY OF SOME
 CC NUCLEAR RECEPTORS BY PROMOTING CHROMATIN CONDENSATION, THUS
 CC PREVENTING ACCESS OF THE BASAL TRANSCRIPTION.
 CC -1- SUBUNIT: FORMS A LARGE COREPRESSOR COMPLEX THAT CONTAINS SIN3A/B
 CC AND HISTONE DEACETYLASES HDAC1 AND HDAC2. THIS COMPLEX ASSOCIATES
 CC WITH THE THYROID (TR) AND THE RETINOIC ACID RECEPTOR (RAR) IN THE
 CC ABSENCE OF LIGAND, AND MAY STABILIZE THEIR INTERACTION WITH TFIIIB.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; SMRT/TRAC-2 (SHOWN HERE) AND
 CC TRAC-1. ARE PRODUCED BY ALTERNATIVE SPLICING. TRAC-1 CONTAINS ONLY
 CC THE C-TERMINAL RECEPTOR-INTERACTING DOMAIN AND ACTS AS AN
 CC ANTIREPRESSOR.
 CC -1- TISSUE SPECIFICITY: UBIQUITOUS. HIGH LEVELS OF EXPRESSION ARE
 CC DETECTED IN LUNG, SPLEEN AND BRAIN.
 CC -1- INDUCTION: REGULATED DURING CELL CYCLE PROGRESSION.
 CC -1- DOMAIN: THE N-TERMINAL REGION CONTAINS REPRESSION FUNCTIONS THAT
 CC ARE DIVIDED INTO THREE INDEPENDANT REPRESSION DOMAINS (RD1, RD2
 CC AND RD3). THE C-TERMINAL REGION CONTAINS THE NUCLEAR RECEPTOR-
 CC INTERACTING DOMAINS THAT ARE DIVIDED IN TWO SEPARATE INTERACTION
 CC DOMAINS (ID1 AND ID2).
 CC -1- DOMAIN: THE TWO INTERACTION DOMAINS (ID) CONTAIN A CONSERVED
 CC SEQUENCE REFERRED TO AS THE CORNR BOX. THIS MOTIF IS REQUIRED AND
 CC SUFFICIENT TO PERMIT BINDING TO UNLIGANDED TR AND RARS. SEQUENCES
 CC FLANKING THE CORNR BOX DETERMINE NUCLEAR HORMONE RECEPTOR
 CC SPECIFICITY.
 CC -1- SIMILARITY: CONTAINS 1 SANT-A DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 MYB-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 2 CORNR BOXES.
 CC -1- SIMILARITY: BELONGS TO THE N-COR NUCLEAR RECEPTOR COREPRESSORS
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF113003; AAD20946.1; -;
 DR EMBL: AF125672; AAD22973.1; -;
 DR EMBL: U37146; AAC50236.1; -;
 DR EMBL: S83390; AAB50847.1; -;
 DR EMBL: U80750; AAB91446.1; -;
 DR TRANSFAC: T04689; -;
 DR Genew: HGNC:7673; NCOR2.


```

FT DOMAIN 466 574 ARG/SER-RICH.
FT DOMAIN 664 814 ARG/SER-RICH.
FT DOMAIN 1311 1348 ARG-SER-TANDEM REPEAT-RICH.
SQ SEQUENCE 1462 AA; 165676 MW; D98A1147763EF527 CRC64;

Query Match
Best Local Similarity 28.1%; DB 1; Length 1462;
Matches 32; Conservative 14; Mismatches 41; Indels 27; Gaps 4;

QY 78 YKD---ASTIRKNSRLSRLDIYKVCNFCGAYFVLENPNPRDIKFDDEPHKRRKRSK 134
DB 681 YSDRSSESPRSKRSRSRSYSTR-----SRSLASSHSRSR 720
QY 135 SOSKSSOTRN---NRSOSNANAHTSKKRDTRKROERHIKEEQDKEMTSKOH 185
DB 721 SPSSRSNSKYSKSDHSCSSSYTIS-SDGGRAKRLRSRSGKNSVSHKKH 773

RESULT 14
SFR6.HUMAN
ID SFR6.HUMAN STANDARD; PRT: 344 AA.
AC Q13247; Q13244; Q13245; Q9UJB8; Q96J06;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Splicing factor, arginine/serine-rich 6 (Pre-mRNA splicing factor
DE SRP55)
GN SFR56 OR SRP55.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RC SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RP TISSUE=Colon;
RX MEDLINE=96016206; PubMed=7556075;
RA Screation G.R., Caceres J.F., Mayeda A., Bell M.V., Plebanski M.,
RA Jackson D.G., Bell J.I., Krainer A.R.;
RT "Identification and characterization of three members of the human SR
RT family of pre-mRNA splicing factors.";
RL EMBO J. 14:4336-4349(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21638749; PubMed=11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baggeley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Copley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Graham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Levasiaiho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConachie L.J., Mcclay K., McMurray A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkneen R., Sims S.,
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A.C., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RT "The DNA sequence and comparative analysis of human chromosome 20.";
RL Nature 414:865-871(2001).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM SRP55-1).

```

```

RC TISSUE=Placenta;
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE ON 21-27 AND 47-55.
RX MEDLINE=92:249775; PubMed=1577277;
RA Zahler A.M., Lane W.S., Stolk J.A., Roth M.B.;
RT "SR proteins: a conserved family of pre-mRNA splicing factors.";
RL Genes Dev. 16:837-847(1992).
CC -!- FUNCTION: PLAYS A ROLE IN CONSTITUTIVE SPLICING AND CAN MODULATE
CC THE SELECTION OF ALTERNATIVE SPLICE SITES.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; SRP55-1 (SHOWN HERE), SRP55-
CC 2 AND SRP55-3; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- PTM: EXTENSIVELY PHOSPHORYLATED ON SERINE RESIDUES IN THE RS
CC DOMAIN (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).
CC -!- SIMILARITY: BELONGS TO THE SR FAMILY OF SPLICING FACTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL: U30833; AAA93073.1; .
DR EMBL: U30828; AAA93071.1; .
DR EMBL: U30829; AAA93072.1; .
DR EMBL: AL031681; CAB43960.1; .
DR EMBL: BC005832; AAB06832.1; .
DR Genew: HGNC:10788; SFR56.
DR MIM: 601944; .
DR InterPro: IPR000504; RNA_rec_mot.
DR Pfam: PF00076; rrm; 2.
DR SMART: SM00360; RRM; 2.
DR PROSITE: PS01102; RRM; 2.
DR PROSITE: PS00030; RRM_RNP_1; FALSE_NEG.
KW Nuclear protein; RNA-binding; mRNA splicing; Alternative splicing;
KW Repeat; Phosphorylation.
FT DOMAIN 1 72
FT DOMAIN 87 90 RNA-BINDING (RRM) 1.
FT DOMAIN 110 183 RNA-BINDING (RRM) 2.
FT DOMAIN 184 343 ARG/SER-RICH (RS DOMAIN).
FT VARSPPLIC 86 135 SGGGYSRRSPGRDKYGPVPTERTYLIVLENLSSRCSWQDL
FT KDFMRQAGE -> MTNGAEAVSTEAKMTAFDPWPLFHTLC
FT DPCPMTLTLTPEAMTTAATFCH (IN ISOFORM SRP55-
FT 2).
FT VARSPPLIC 136 344 MISSING (IN ISOFORM SRP55-2).
FT VARSPPLIC 313 344 RSVPPPKRATSRSRSRKSRSRSSSRD -> LKLGA
FT CONFLICT 64 64 RFMSQQTESIYSLASSC (IN ISOFORM SRP55-3).
FT SEQUENCE 344 AA; 39586 MW; 72305506CE948B94 CRC64;
SQ
Query Match
Best Local Similarity 23.1%; DB 1; Length 344;
Matches 52; Conservative 34; Mismatches 83; Indels 56; Gaps 9;

QY 17 GPQFTIFENIVKSDDIEFQVLVINEKSAFDTV---FGQRFSEILLKYDFIVGDWNEQLR 73
DB 36 GYGFEFEDSRDADDAVVELNGKELCGERVIVEHARGPRDRDGYSGRSGGGYSSRR 95
QY 74 LRGFVKDASTIKNSRI-----SR-----LEDYI-----KEYCNFGCAYF-- 108
DB 96 TSGRIKYGPPVTEYRLIVLENLSSRCSWQDLKDFMRQAGEVTYADAHKERTNEGVEFRS 155
QY 109 -----VLE-----NPNPRDIKFDDEP-----HKRRKSRKS-QSSKSO 141
DB 156 YSDMRALDKLDGTEINGNRNIKIEDKPTSHRSYSGSRSRSRSRSRKSRSR 215
QY 142 TRN-IRSQSNANAHTSKKRDTRKROERHIKEEQDKEMTSKOH 185
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM SRP55-1).

```


[3] X-RAY CRYSTALLOGRAPHY (2.15 ÅNGSTROMS) OF 1-177.
RN RP STRAIN=ATCC 75593;
RC
RX MEDLINE=98318040; PubMed=9655334;
RA Mikol V., Ma D., Carlow C.K.S.;
RX "Crystal structure of the cyclophilin-like domain from the parasitic
RT nematode *Brugia malayi*.";
RL Protein Sci. 7:1310-1316(1998).
RN [4]
RN RP X-RAY CRYSTALLOGRAPHY (2.47 ÅNGSTROMS) OF 1-177.
RX MEDLINE=20108543; PubMed=10642184;
RC Ellis P.J., Carlow C.K.S., Ma D., Kuhn P.;
RX "Crystal structure of the complex of *Brugia malayi* cyclophilin and
RT

RL Biochemistry 39:592-598(2000).
CC -|- FUNCTION: PPIases accelerate the folding of proteins. It catalyzes
CC the cis-trans isomerization of proline imidic peptide bonds in
CC oligopeptides.
CC
CC -|- CATALYTIC ACTIVITY: Peptidylproline (omega=180) = peptidylproline
CC (omega=0).
CC
CC -|- ENZYME REGULATION: RELATIVELY INSENSITIVE TO INHIBITION BY CSA.
CC
CC -|- SIMILARITY: BELONGS TO THE CYCLOPHILIN-TYPE PPIASE FAMILY.
CC
CC -----
CC

CC This SWISS-PROT entry is copyright, it is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L37292; AAC37249.1; -.
 CC PDB; 1A58; 27-MAY-98.
 CC PDB; 1A33; 29-JUL-98.
 CC PDB; 1C5F; 02-DEC-99.
 CC InterPro; IPR002130; CSA_Pptase.
 CC Pfam; PF00160; pro_isomerase; 1.
 CC PRINTS; PR00153; CSAPPISMRAE.
 CC PROSITE; PS00170; CSA_PPTASE_1; 1.
 CC PROSITE; PSS0072; CSA_PPTASE_2; 1.
 CC Isomerase; Rotamase; 3d-structure.
 CC 10 175 PPTASE; CYCLOPHILIN-TYPE.
 CC 700 709 POLY-ARG.
 CC DOMAIN
 CC FT
 CC FT

```

FT      DOMAIN          800      815      POLY-SER.
FT      DOMAIN          828      837      POLY-SER.
SQ      SEQUENCE        843 AA; 97817 MW; 3C34EC90A32EDBDC CRC64;

Query Match           9.6%; Score 96.5; DB 1; Length 843;
Best Local Similarity 21.9%; Pred. No. 1.7;
Matches 41; Conservative 37; Mismatches 56; Indels 53; Gaps

Qy    28   KSDDIEFQLVINEKSAFDVTIVFGORFSE-ILLKYDFIVGDWGNQEOLRLRGFYKDASTIR 85
       | ||| : : : ||| | : : : | : : : | : : : | : : : | : : : | : : : |
Db    603  KLDDLKRKETSGOKSQAD-----SEQVEAKTNVVDNSDNKSVMNGKLKEVSSTN 654
               ||| | : : : | : : : | : : : | : : : | : : : | : : : |
Qy    86   KNSRISLRELY-----IKEYCNFGCAFYVLE-----NPNRPDIKFDDDERPHKRR 129
       | : : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db    655   KENEVSEQDKLKAEETKSEELKQOVN-----EVSRSKQGKGEPKHEHKNERSRSRRR 706
               ||| | : : : | : : : | : : : | : : : | : : : | : : : |
Qy    130  KS-----RSKQSOSKTSOTNNRSQSNANAHT-----SKKKRDTKR-----RQERH 170
       : | : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db    707  RSRNGRRRRSSRRSRDRRHKSRSRGYVRFFEGWSRSRPTRELRYDMMRRERE 766
               ||| | : : : | : : : | : : : | : : : | : : : | : : : |
Qy    171  IKEEQDK 177
       : : :
Db    767  RRRSFRD 773
               ;

RESULT 10
APU_THESA
ID     APU_THESA
AC     P36905;

```

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>), or send an email to license@isb-sib.ch).

```

CC DR EMBL: M74002; AAA35554.1; -.
CC DR Genbank; HGNC:10782; SFRS11.
CC DR MIM; 602010; -.
CC DR InterPro; IPR000504; RNA_rec_mot.
CC DR Pfam; PF00076; rrm; 1.
CC DR SMART; SM00360; RRM; 1.
CC DR PROSITE; PS50102; RRM; 1.
CC DR PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
KW Nuclear protein; RNA-binding; mRNA splicing; Repeat.
FT DOMAIN 17 31
FT DOMAIN 33 113
FT DOMAIN 252 259
FT DOMAIN 291 294
FT DOMAIN 341 349
FT DOMAIN 397 404
FT DOMAIN 247 353
FT REPEAT 247 255
FT REPEAT 258 265
FT REPEAT 267 274
FT REPEAT 275 282
FT REPEAT 285 292
FT REPEAT 293 300
FT REPEAT 302 309
FT REPEAT 321 328
FT REPEAT 334 341
FT REPEAT 346 353
SQ SEQUENCE 484 AA; 53542 MW; 1211E96DDCOA3182 CRC64;

Query Match          9.8%; Score 99; DB 1; Length 484;
Best Local Similarity 34.3%; Pred. No. 0.57;
Matches 24; Conservative 17; Mismatches 21; Indels 8; Gaps 2;

Oy 122 DERPHKRKRSKQS-----SKSQTNRNRSQSNANAHFTSKRKDKTK--RRQERHIKE 173
Db 240 DKKEKRHRHSRKSRRRRTPSSSRHRRSRSSRRSRSSRRSRSSRRSRSSRRSRSS 299
Oy 174 EDDKEMTSK 183
Db 300 RGRSRSTSK 309

RESULT 7
SFR5_MOUSE
ID SFR5_MOUSE STANDARD; PRT; 270 AA.
AC O35326;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Splicing factor, arginine/serine-rich 5 (Pre-mRNA splicing factor SRP40) (Delayed-early protein HRS).
GN SFR5 OR HRS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=98094279; PubMed=9434190;
RA Du K., Taub R.;
RT "Alternative splicing and structure of the human and mouse SFR5/HRS/SRP40 genes.";
RL Gene 204:243-249(1997).
CC -!- FUNCTION: MAY BE REQUIRED FOR PROGRESSION THROUGH G1 AND ENTRY INTO S PHASE OF CELL GROWTH. MAY PLAY A REGULATORY ROLE IN PRE-MRNA SPLICING. AUTOREGULATES ITS OWN EXPRESSION. PLAYS A ROLE IN CONSTITUTIVE SPLICING AND CAN MODULATE THE SELECTION OF ALTERNATIVE SPLICING SITES (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).

```

```

CC CC -!- PTM: EXTENSIVELY PHOSPHORYLATED ON SERINE RESIDUES IN THE RS DOMAIN (BY SIMILARITY).
CC CC -!- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).
CC CC -!- SIMILARITY: BELONGS TO THE SR FAMILY OF SPLICING FACTORS.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC CC -----
CC DR EMBL: AF020308; AAC39946.1; -.
CC DR MGD; MGI:98287; Sfrs5.
CC DR InterPro; IPR000504; RNA_rec_mot.
CC DR Pfam; PF00076; rrm; 2.
CC DR SMART; SM00360; RRM; 2.
CC DR PROSITE; PS50102; RRM; 2.
CC DR PROSITE; PS00030; RRM_RNP_1; 1.
KW Transcripton regulation; Nuclear protein; RNA-binding; mRNA splicing; Repeat; Phosphorylation.
FT DOMAIN 4 74
FT DOMAIN 78 83
FT DOMAIN 108 190
FT DOMAIN 191 265
FT DOMAIN 270 AA; 30944 MW; 277397C3DC657895 CRC64;
SQ SEQUENCE 270 AA; 30944 MW; 277397C3DC657895 CRC64;

Query Match          9.8%; Score 98.5; DB 1; Length 270;
Best Local Similarity 20.9%; Pred. No. 0.33;
Matches 44; Conservative 46; Mismatches 68; Indels 53; Gaps 8;

Oy 20 FIHEHIVKSDDIEFQVLVINEKSAFDVTIV-----FCQ-RPEILLKYDFIVGQWG 63
Db 41 FVEFEDPRDADAVYELDGKELCSERVITIEHARARSGRGRGYSDDL--G--AVADLG 95
Oy 69 NEQLR-----LRGFYKDASTIR-KNSRISRLDYIKEYCNFG 104
Db 96 NDRKKCSTCKKRNRLIVENLSRVSWQDLKDFMQAGEVTFADAHKPLKNGVVEFAFYG 155
Oy 105 CAYEVLENPNPRDIKFDDEP-----HKRRKSRSKQS-----SKSQTNRNRSQSNAN 152
Db 156 DLKNAIEKLSWKR-KLTGKSKLIEGSKRHSRHSRHSRHSRHSRHSRHSRHSRHSRHSR 214
Oy 153 AHFTS(KRKDTKRQERHIKEEQDKEMTSK 183
Db 215 SRSRSRSRSKRSRSGSRSPVPEKSKQKRGSSSR 245

RESULT 8
RA50_SULAC
ID RA50_SULAC STANDARD; PRT; 886 AA.
AC O33600;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA double-strand break repair rad50 ATPase.
GN RAD50.
OS Sulfolobus acidocaldarius.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OX NCBI_TaxID=2285;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33909 / NCIB 11770 / DSM 639;
RX MEDLINE=97362314; PubMed=9211741;
RA Elie C., Baucher M.F., Fondrat C., Forterre P.;
RT "A protein related to eucaryal and bacterial DNA-motor proteins in the hyperthermophilic archaeon Sulfolobus acidocaldarius.";
RL J. Mol. Evol. 45:107-114(1997).
CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The rad50/mre11 complex possesses single-strand endonuclease activity and ATP-dependent double-strand-specific exonuclease activity.
CC CC -----

```


QY 156 TSKKRKDTKRQERHKEQDKEMTSK 183
 ID SFRS_HUMAN STANDARD; PRT; 272 AA.
 AC Q13243; Q16662; O14797;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Splicing factor, arginine/serine-rich 5 (pre-mRNA splicing factor
 DE SRP40) (Delayed-early protein HRS).
 GN SFRS5 OR SRP40 OR HRS.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Colon;
 RX MEDLINE=96016206; PubMed=7556075;
 RA Sreaton G.R., Caceres J.F., Mayeda A., Bell M.V., Plebanski M.,
 RA Jackson D.G., Bell J.R., Krainer A.R.;
 RT "Identification and characterization of three members of the human SR
 RT family of pre-mRNA splicing factors.";
 RL EMBO J. 14:4336-4349(1995).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Skin;
 RA Strausberg R.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1-192 FROM N.A.
 RC TISSUE=Thymus;
 RX MEDLINE=98094279; PubMed=9434190;
 RA Du K., Taub R.;
 RT "Alternative splicing and structure of the human and mouse
 RT SFRS5/HRS/SRP40 genes.";
 RL Gene 204:243-249(1997).
 RN [4]
 RP SEQUENCE OF 130-158.
 RX MEDLINE=92249775; PubMed=1577277;
 RA Zahler A.M., Lane W.S., Stolk J.A., Roth M.B.;
 RT "SR proteins: a conserved family of pre-mRNA splicing factors.";
 RL Genes Dev. 6:837-847(1992).
 CC -1- FUNCTION: PLAYS A ROLE IN CONSTITUTIVE SPLICING AND CAN MODULATE
 CC THE SELECTION OF ALTERNATIVE SPLICE SITES.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- ALTERNATIVE PRODUCTS: 4 ISOFORMS; SRP40-1 (SHOWN HERE), SRP40-
 CC 2, SRP40-3 AND SRP40-4; ARE PRODUCED BY ALTERNATIVE SPLICING AND
 CC DIFFER IN THEIR C-TERMINAL DOMAINS.
 CC -1- PTM: EXTENSIVELY PHOSPHORYLATED ON SERINE RESIDUES IN THE RS
 CC DOMAIN (BY SIMILARITY).
 CC -1- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).
 CC -1- SIMILARITY: BELONGS TO THE SR FAMILY OF SPLICING FACTORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U30826; AAA93070.1;
 CC EMBL; U30884; AAA93074.1;
 CC EMBL; U30827; AAB60405.1;
 CC EMBL; BC018823; AAU18823.1;
 CC EMBL; AF020307; AAC39543.1;
 CC Genbank; HGNC:10787; SFRS5.

DR MIM; 600914;
 DR InterPro: IPR000504; RNA_rec_mot.
 DR Pfam; PF00076; rrm; 2.
 DR SMART; SM00360; RRM; 2.
 DR PROSITE; PS0102; RRM; 2.
 DR PROSITE; PS00030; RRM_RNP_1; 1.
 KW Nuclear protein; RNA-binding; mRNA splicing; Alternative splicing;
 KW Repeat; Phosphorylation.
 FT DOMAIN 4 74 RNA-BINDING (RRM) 1.
 FT DOMAIN 78 83 GLY-RICH (HINGE REGION).
 FT DOMAIN 108 181 RNA-BINDING (RRM) 2.
 FT DOMAIN 182 267 ARG/SER-RICH (RS DOMAIN).
 FT VARSPLIC 43 45 MISSING (IN ISOFORM SRP40-4).
 FT VARSPLIC 100 107 NAPPVTE -> YVKGGLWLH (IN ISOFORM SRP40-
 FT 2).
 FT VARSPLIC 108 272 MISSING (IN ISOFORM SRP40-2).
 FT VARSPLIC 113 140 ENLSSRVSWQDLKDFMRQAGVTFADAH -> GEFTLKSQL
 FT AGSQRFHETSWGNSVCGCT (IN ISOFORM SRP40-4).
 SQ SEQUENCE 272 AA; 31263 MW; F13AD79845ECB16 CRC64;
 Query Match 9.9%; Score 99.5; DB 1; Length 272;
 Best Local Similarity 20.4%; Pred. No. 0.28;
 Matches 43; Conservative 44; Mismatches 73; Indels 51; Gaps 8;
 QY 20 FIIFENIVKSDIEFQVLVINEKSAFDVTV-----FGQFPSEILLKYD--- 61
 Db 41 FVEFEDPRDADAVYELDGKELCSERVIEHARARSGRGRGYSRDFSSRRPRNDRN 100
 QY 62 -----FIVGD-----WGNEQLRLRGFYKDASTIR-KNSRISLEDYIKEYCNFGC 105
 Db 101 APPVPTENRLIVENLSRVSWQD-----LKDFMRQAGVTFADAHRLKLVGVVEFASIGD 156
 QY 106 AYFVLE-----NPNPRDIK-FDDERPHKRKRSKSKS-----SKSQTNRNNSQSNAN 152
 Db 157 LKNAIEKLSCGKEINGRKIKLEGRSHRSKRSRSTRSSRSRSTRSSRSRSTRSSRSRSTRSSRSR 216
 QY 153 AHFTSKRKDTKRQERHKEQDKEMTSK 183
 Db 217 SRSRSRSKRSKRSVSRSPVPEKSQKRGSSSR 247
 RESULT 6
 SFRB_HUMAN STANDARD; PRT; 484 AA.
 ID SFRB_HUMAN STANDARD; PRT; 484 AA.
 AC Q05519;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Splicing factor arginine/serine-rich 11 (Arginine-rich 54 kDa nuclear
 DE protein) (p54).
 GN SFRS11.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Hepatoma;
 RX MEDLINE=91376109; PubMed=1896467;
 RA Chaudhary N., McMahon C., Blobel G.;
 RT "Primary structure of a human arginine-rich nuclear protein that
 RT colocalizes with spliceosome components.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:8189-8193(1991).
 CC -1- FUNCTION: MAY FUNCTION IN PRE-MRNA SPLICING.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR. COLOCALIZES WITH SPliceosome
 CC COMPONENTS.
 CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

```
CC -----
DR EMBL; L14076; AAA36649.1; -.
DR EMBL; BC002781; AA020781.1; -.
DR EMBL; AC004236; AAC04476.1; -.
DR Genew; HGNC:10786; SFRS4.
DR MIM; 601940; -.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 2.
DR SMART; SM00360; RRM; 2.
DR PROSITE; PS00102; RRM; 2.
DR PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
KW mRNA processing; mRNA splicing; Nuclear protein; RNA-binding;
KW Repeat; Phosphorylation.
FT DOMAIN 2 72 RNA-BINDING (RRM) 1.
FT DOMAIN 72 97 GLY-RICH (HINGE REGION).
FT DOMAIN 104 177 RNA-BINDING (RRM) 2.
FT DOMAIN 179 494 ARG/SER-RICH (RS DOMAIN).
FT CONFLICT 35 35 N -> D (IN REF. 4).
FT CONFLICT 253 253 E -> D (IN REF. 1).
FT CONFLICT 318 322 SRGRS -> EQQEE (IN REF. 1).
FT CONFLICT 338 338 G -> A (IN REF. 1).
FT CONFLICT 356 356 G -> S (IN REF. 1).
FT CONFLICT 436 438 TNQ -> RNE (IN REF. 1).
SQ SEQUENCE 494 AA; 56678 MW; 5BBAB917C218C20A CRC64;

Query Match 10.1%; Score 102; DB 1; Length 494;
Best Local Similarity 22.4%; Pred. No. 0.34;
Matches 52; Conservative 35; Mismatches 75; Indels 70; Gaps 10;

Qy 17 GPQTFHFNIVKSDIEFOLVINEKSAFDVTV-----FGQRFS----- 54
Db 36 GYGFVEFDLDDADDAVYELNGKDLGGERVIVEHARGPRDGSYGSGRGYRRSGRDK 95
Qy 55 ---EILLXYDFIVGD-----WGN--EQLRUG--FYKDASTIRKNSIRLEDYIKEYC 101
Db 96 YGPPTRTYRLIVENLSSRVSWQD---LKDFMROAGEVTFADAHKPKLNEGVEEASG 148
Qy 102 NFGCAYFYLE-----NPNPRDIKFDDEP-----HKR-----RKSRKS 135
Db 149 SYSOMKRALEKLDGTEVNGRKIRLVEDKPGRRRSYSRSHSRSHSRSHSRSR 208
Qy 136 QSSKSTQNNRSQSNANAHFTSK--KRKDTKRQERHKEBODKEMTSAKOH 185
Db 209 GSSKSSHSKRSRSGRSRSGRSRSGRSRSGRSRSGRSRSGRSRSGRSRSH 260

RESULT 4
SFR5_RAT STANDARD: PRT; 269 AA.
AC Q09167; Q35335;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Splicing factor, arginine/serine-rich 5 (Pre-mRNA splicing factor
DE SRP40) (Insulin-induced growth response protein CL-4) (Delayed-early
DE protein HRS).
GN SFRS5 OR CL-4 OR HRS.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA MEDLINE=93315501; PubMed=7686911;
RA Diamond R.H., Du K., Lee V.M., Mohr K.L., Haber B.A., Tewari D.S.,
RA Taub R.;
RT "Novel delayed-early and highly insulin-induced growth response
RT genes. Identification of HRS, a potential regulator of alternative
RT pre-mRNA splicing."
RL J. Biol. Chem. 268:15195-15192(1993).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
```

```
RC TISSUE=Thymus;
RX MEDLINE=98094279; PubMed=9434190;
RA Du K., Taub R.;
RT "Alternative splicing and structure of the human and mouse
RT SFRS5/HRS/SRP40 genes."
RL Gene 204:243-249(1997).
RN [3]
RP SEQUENCE OF 185-269 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Testis;
RX MEDLINE=94164020; PubMed=8161377;
RA Hamil K.G., Hall S.H.;
RT "Cloning of rat Sertoli cell follicle-stimulating hormone primary
RT response complementary deoxyribonucleic acid: regulation of TSC-22
RT gene expression."
RL Endocrinology 134:1205-1212(1994).
CC -1- FUNCTION: MAY BE REQUIRED FOR PROGRESSION THROUGH G1 AND ENTRY
CC INTO S PHASE OF CELL GROWTH. MAY PLAY A REGULATORY ROLE IN PRE-
CC MRNA SPLICING. AUTOREGULATES ITS OWN EXPRESSION. PLAYS A ROLE IN
CC CONSTITUTIVE SPLICING AND CAN MODULATE THE SELECTION OF
CC ALTERNATIVE SPICE SITES (BY SIMILARITY).
CC -1- FUNCTION: COULD PLAY AN IMPORTANT ROLE IN DEVELOPMENT AND
CC DIFFERENTIATION IN THE SPLEEN AND THYMUS.
CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2/HRR-LF; are
CC produced by alternative splicing.
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN SPLEEN AND THYMUS.
CC -1- INDUCTION: BY INSULIN AND HEPATECTOMY.
CC -1- PTM: EXTENSIVELY PHOSPHORYLATED ON SERINE RESIDUES IN THE RS
CC DOMAIN (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).
CC -1- SIMILARITY: BELONGS TO THE SR FAMILY OF SPLICING FACTORS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L13615; AAA62266.1; -.
DR EMBL; AF020683; AAB1864.1; -.
DR EMBL; L33267; AAA42316.1; -.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 2.
DR SMART; SM00360; RRM; 2.
DR PROSITE; PS00102; RRM; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
DR Transcription regulation; Nuclear protein; RNA-binding; mRNA splicing;
KW Repeat; Phosphorylation; Alternative splicing.
FT DOMAIN 4 74 RNA-BINDING (RRM) 1.
FT DOMAIN 75 107 ARG/GLY-RICH (HINGE REGION).
FT DOMAIN 108 181 RNA-BINDING (RRM) 2.
FT DOMAIN 184 264 ARG/SER-RICH (RS DOMAIN).
FT VARSPIC 123 124 DL -> VC (IN ISOFORM 2).
FT VARSPIC 125 269 MISSING (IN ISOFORM 2).
FT CONFLICT 69 69 I -> M (IN REF. 2).
SQ SEQUENCE 1269 AA; 30891 MW; 0FC90628B64DE845 CRC64;

Query Match 9.9%; Score 100; DB 1; Length 269;
Best Local Similarity 20.7%; Pred. No. 0.25;
Matches 43; Conservative 43; Mismatches 74; Indels 48; Gaps 7;

Qy 20 FIHEENIVKSDIEFOLVINEKSAFDVTV-----FGQRFSEILLKYD--- 61
Db 41 FVEFEDPDADDAVYELDGKELCSERVITIEHARSRGGRGGRYSRDRFSRRPRNDRN 100
Qy 62 -----FTVGD-----WGNQLRLRGYKDASTR-KNSRISRLEDYIKEYCNGFC 105
Db 101 APPVTEENRLIVENLSSRVSWQD---LKDFMROAGEVTFADAHKPKLNEGVEEASG 156
Qy 106 AYFVLE-----NPNPRDIKFDDEPHKRSKRSQS-----SKSQTNNRSQSNANAH 155
Db 157 LKNAIEKLSGKINGRKIKLEGSRKRSRSTRSSRSRSGRSRSGRSRSGRSR 216
```



```
Db 1128 SNKKEDKNEKKKSHQHVLYKK 1149

RESULT 39
Q8REF7
AC Q8REF7 PRELIMINARY; PRT; 1056 AA.
DT 01-JUN-2002 (TrEMBLrel. 21, Created).
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE ATP-dependent nuclease subunit A.
GN FN1149.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RC SPRAIN=ATCC 25586;
RX MEDLINE=21886394; PubMed=11889109;
RA Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A.,
RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
RA Fonstein M., Kyripides N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
RT nucleatum strain ATCC 25586.";
RL J. Bacteriol. 184:2005-2018(2002).
DR EMBL; AE010620; AAL95345.1; -.
KW Complete proteome.
SQ SEQUENCE 1056 AA; 125480 MW; D07C14C2AC6F6465 CRC64;

Query Match 9.0%; Score 90.5; DB 16; Length 1056;
Best Local Similarity 22.5%; Pred. No. 27;
Matches 48; Conservative 41; Mismatches 67; Indels 57; Gaps 11;

QY 8 EMLNKKYGPQF-----IHFENIVKSDIEFQVINEKSAFDVTF-----GQR--- 52
Db 736 ELVTIHKSGLEKFTTVINKSKKSTDDIDFLFEMNDK--YDKTVSFLCKGKYPILK 793
QY 53 --FSEILLKYDFIVGDMGNQLRGYKDASTRKNSRISLEDYIKCYNFGCAVFL 110
Db 794 TCPERIENYDKI-----KEEENINFYV-ALTRPKNNLIVIEDRLF----- 836
QY 111 ENP-----NPRD-----TKFDDERPHKRRKRSK---SQSKSQTRNNRSQSNANAHF 155
Db 837 ENPLNESNIDDFNCELGKISLDE-----KSKTEDIIEKNLENDLYNSQSYFSSSIYE 890
QY 156 TSKRKDTKRRQERHIKEEODKEMTSAKQHLF 188
Db 891 NEEEKNIEVNESKFLLETEEKRMIGILVHYFF 923

RESULT 40
Q9UHV3
AC Q9UHV3 PRELIMINARY; PRT; 1103 AA.
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Pancreas-enriched phospholipase C.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kawasaki H., Chen E.J., Springett G.M., Graybiel A.M., Housman D.E.;
RT "A novel phospholipase C enriched in pancreas.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF117948; AAF22005.1; -.
DR HSSP; P10688; 1DJX.
DR InterPro; IPR000008; C2.
DR InterPro; IPR001192; PI-PLC.
```

```
DR InterPro; IPR000909; PI-PLC_xdom.
DR InterPro; IPR001711; PI-PLC_Y.
DR InterPro; IPR000159; RA_domain.
DR Pfam; PF00168; C2; 1.
DR Pfam; PF00388; PI-PLC-X; 1.
DR Pfam; PF00387; PI-PLC-Y; 1.
DR Pfam; PF00788; RA; 1.
DR PRINTS; PR00390; PHPLIPASEC.
DR ProDom; PD001202; PI-PLC_Y; 1.
DR SMART; SM00239; C2; 1.
DR SMART; SM00148; PLCXC; 1.
DR SMART; SM00149; PLCYC; 1.
DR SMART; SM00314; RA; 1.
DR PROSITE; PS50004; C2_DOMAIN_2; 1.
DR PROSITE; PS50007; PIPLC_X_DOMAIN; 1.
DR PROSITE; PS50008; PIPLC_Y_DOMAIN; 1.
SQ SEQUENCE 1103 AA; 124905 MW; 29BC730DB4377995 CRC64;

Query Match 9.0%; Score 90.5; DB 4; Length 1103;
Best Local Similarity 24.0%; Pred. No. 28;
Matches 46; Conservative 21; Mismatches 62; Indels 63; Gaps 7;

QY 24 ENIVKSDDIEFQVINEKSAFDVTFVFGQRFSEILLKYDFIVGDMGNQLRGYKDAST 83
Db 920 KNIVQ-DDKEVILSSEESFF-----VQVHDVSPQPR-----T 952
QY 84 IRKNSRISLEDYIKE-YCNFGCAVYLENPNRPDIKEDDE----RPHKRRKRSKQ-- 136
Db 953 VIRAPRVSTAQDVIOQTLCRAKYSYLSNPNPSDYVLLLEVVKDNTTKKTTTPKSSQRV 1012
QY 137 -----SSKSQTRNNRSQSNANAHFTSKRKDTKRRQERHIKEE 174
Db 1013 LLDQECVFQAQSKWKAGKGFILKKEQVQASREDKKGISFASSELKLTG-----STK 1065
QY 175 QDKEMTSAKOHL 186
Db 1066 QPRGLTSPSQLL 1077

Search completed: February 26, 2003, 09:54:18
Job time : 36 secs
```


US Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;


```
Q9DBP1
ID Q9DBP1 PRELIMINARY; PRT; 339 AA.
AC Q9DBP1;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE 1210001ELIRIK protein.
GN 1210001ELIRIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LUNG;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Bojelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK004831; BAB23599.1; -.
DR MGD: MGI:1915246; 1210001ELIRIK.
DR InterPro: IPR000504; RNA_rec_mot.
DR Pfam: PF00076; rrm; 2.
DR SMART: SM00360; RRM; 2.
DR PROSITE: PS0102; RRM; 2.
SQ SEQUENCE 339 AA; 38997 MW; BC3C9BDA38144A0D CRC64;

Query Match 9.3%; Score 94; DB 11; Length 339;
Best Local Similarity 23.3%; Pred. No. 4;
Matches 50; Conservative 29; Mismatches 88; Indels 48; Gaps 7;

Qy 17 GPQFIHFENIVKSDIEFOLVINEKSAFDVTV---FGQFSEILLKYDFIVGDNGNEQLR 73
Db 36 GYGVEFEEDSRDADAVYELNSKELCGERVIVEHARGPRDRDGYSGSRGGGYSSRR 95

Qy 74 LRGFYKDASTIRKNSRI-----SR-----LEDYI-----KEYCNFGCAVFF-- 108
Db 96 TSGRDYKGGPPVRYEYRLIVENLSRCSQWDLKDFMRQAGEVYADAHKERTNEGVIERS 155

Qy 109 -----VLE-----NPNPRDIKFDPERH-----KRRKRSKSSQSSKSTNNRSQ 148
Db 156 YSDMKRALDKLDGTEINGRNIRLIEDKPTSHRRSYSGSRSRSRKRSRSRSSRRSR 215

Qy 149 SNANAHTFSKKRKTQRQERHIKEQDKEMTSK 183
Db 216 SRISIKSRSRSRKSRKSRKSRKSRKSRKSK 250

RESULT 29
Q9LQ35
ID Q9LQ35 PRELIMINARY; PRT; 462 AA.
AC Q9LQ35;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE F14M2.4 protein.

Qy 17 GPQFIHFENIVKSDIEFOLVINEKSAFDVTV---FGQFSEILLKYDFIVGDNGNEQLR 73
Db 36 GYGVEFEEDSRDADAVYELNSKELCGERVIVEHARGPRDRDGYSGSRGGGYSSRR 95

Qy 74 LRGFYKDASTIRKNSRI-----SR-----LEDYI-----KEYCNFGCAVFF-- 108
Db 96 TSGRDYKGGPPVRYEYRLIVENLSRCSQWDLKDFMRQAGEVYADAHKERTNEGVIERS 155

Qy 109 -----VLE-----NPNPRDIKFDPERH-----KRRKRSKSKS-QSSKSK 141
Db 156 YSDMKRALDKLDGTEINGRNIRLIEDKPTSHRRSYSGSRSRSRKRSRSRSSRRSR 215

Qy 142 TRN-NRSQSNANAHTFSKKRKTQRQERHIKEQDKEMTSK 185
```

```
GN F14M2.4.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsais.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
RA Altafi H., Araujo R., Huizar L., Rowley D., Buehler E., Dunn P.,
RA Gonzalez A., Kremetska I., Kim C., Lenz C., Li J., Liu S.,
RA Luros S., Schwartz J., Shinn P., Toriumi M., Vysotskaia V.S.,
RA Walker M., Yu G., Ecker J., Theologis A., Davis R.W.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC010164; AAF97280.1; -.
SQ SEQUENCE 462 AA; 53790 MW; 708238ED3C11C3C CRC64;

Query Match 9.3%; Score 94; DB 10; Length 462;
Best Local Similarity 26.6%; Pred. No. 5.6;
Matches 42; Conservative 25; Mismatches 63; Indels 28; Gaps 7;

Qy 38 INEKSAFDVTVFGQFSEILLKYDFIVGDNGNEQLRLRGFYKDASTIRKNSRISRLDYI 97
Db 98 ISQREEFNTLQIFDDKILDY-FIVVFTGGDELE-----ADNQT-----LDYL 142

Qy 98 KEYCNFGCAVFF---VLENPNPRDIKFDPERHKKRSKSKSQT-----RNNRSQSN 150
Db 143 RE-----GCPFLTRVLKCGRKVLNNKTKDKCKRNQNLNLAHVTDIRQNGGIPYT 198

Qy 151 ANAHFTSKKRKTQRQERHIKEQDKEMTSA---KQHL 186
Db 199 ENMHRIKEENDKLREQESNIDAKKLAEEDSAMMKQKL 236

RESULT 30
Q921K3
ID Q921K3 PRELIMINARY; PRT; 339 AA.
AC Q921K3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE RKEN CDNA 1210001E11 gene.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC012039; AAH12039.1; -.
DR InterPro: IPR000504; RNA_rec_mot.
DR Pfam: PF00076; rrm; 2.
DR PROSITE: PS0102; RRM; 2.
SQ SEQUENCE 339 AA; 39029 MW; AE8B6650D0448A0D CRC64;

Query Match 9.2%; Score 93; DB 11; Length 339;
Best Local Similarity 23.1%; Pred. No. 4.9;
Matches 52; Conservative 34; Mismatches 83; Indels 56; Gaps 9;

Qy 17 GPQFIHFENIVKSDIEFOLVINEKSAFDVTV---FGQFSEILLKYDFIVGDNGNEQLR 73
Db 36 GYGVEFEEDSRDADAVYELNSKELCGERVIVEHARGPRDRDGYSGSRGGGYSSRR 95

Qy 74 LRGFYKDASTIRKNSRI-----SR-----LEDYI-----KEYCNFGCAVFF-- 108
Db 96 TSGRDYKGGPPVRYEYRLIVENLSRCSQWDLKDFMRQAGEVYADAHKERTNEGVIERS 155

Qy 109 -----VLE-----NPNPRDIKFDPERH-----KRRKRSKSKS-QSSKSK 141
Db 156 YSDMKRALDKLDGTEINGRNIRLIEDKPTSHRRSYSGSRSRSRKRSRSRSSRRSR 215

Qy 142 TRN-NRSQSNANAHTFSKKRKTQRQERHIKEQDKEMTSK 185
```


QY	61	DFIV---GDW-----GNEQLRLRGFYKDASTIRKNSRISRLDYIKEYCN-----FGCAY	10
Df	71	DFMVGGDGFSEGNRGGEST-YGFPFDESFVKKHNK-----EFLLSMANRGKDTNGSQF	124
QY	108	FVLNPNPR-----DK	119
Df	125	FIUTKPPLDGHVVGVQVISQEVVREINQKTDAASKPFAEVRILSCGELVPKSKVK	184
QY	120	FDDRPKHRRKRKSks----QSSKSQTNRNRSQSNAHAFTSKRKRDTKRQERHKEEQD	176
Df	185	KEEKRRHKSSSSSDSDSDSSDSQSSDSETASEEKSRRKKRKRKSRKHHKKEK	244
QY	177	KEMTSak 183	
Df	245	KRKSKK 251	
RESULT 23			
Q9LW95	ID	Q9LW95 PRELIMINARY; PRT; 513 AA.	
AC	AT	Q9LW95;	
DT	01-OCT-2000	(TReMBLrel. 15, Created)	
DT	01-OCT-2000	(TReMBLrel. 15, Last sequence update)	
DT	01-DEC-2001	(TReMBLrel. 19, Last annotation update)	
DE	KED.		
GN	KED.		
OC	Nicotiana tabacum	(Common tobacco).	
OC	Eukaryota;	Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
OC	Spermatophyta;	Magnoliophyta; eudicotyledons; core eudicots;	
OC	Asteridae;	euastrerids I; Solanales; Solanaceae; Nicotiana.	
OX	NCBI_TaxID=4097;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=LEAF;		
RX	MEDLINE=20399434;	PubMed=10945337;	
RA	Hara K., Yagi M., Koizumi N., Kusano T., Sano H.;		
RT	"Screening of wound-responsive genes identifies an immediate-early		
RT	expressed gene encoding a highly charged protein in mechanically		
RT	wounded tobacco plants.";		
RL	Plant Cell Physiol. 41:684-691(2000).		
DR	EMBL; BA009883; BAA95789.1; -.		
SQ	SEQUENCE 513 AA; 61019 MW; F9598A37A840B970 CRC64;		
Query Match 9.6%; Score 96.5; DB 10; Length 513;			
Best Local Similarity 26.7%; Pred. No. 3.9;			
Matches 20; Conservative 19; Mismatches 25; Indels 11; Gaps			
QY	114	NPRDIKFDDRPHKRV---RKSRKSQSSKSQTRNNRSQSNAHAFTSKRKDTPRQERRH	170
Df	48	NVREKYKYEKKDKEKKESSEETEEKDD-----GKGKKDKKKKKHTD	99
QY	171	IKEEQDKEMTSAKQH 185	
Df	100	MKEKDKEMDKSKH 114	
RESULT 24			
Q99P84	ID	Q99P84 PRELIMINARY; PRT; 2281 AA.	
AC	AT	Q99P84;	
DT	01-JUN-2001	(TReMBLrel. 17, Created)	
DT	01-JUN-2001	(TReMBLrel. 17, Last sequence update)	
DT	01-JUN-2002	(TReMBLrel. 21, Last annotation update)	
DE	Phosphonositide-specific phospholipase C epsilon.		
OS	Rattus norvegicus	(Rat).	
OC	Eukaryota;	Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia;	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
OX	NCBI_TaxID=10116;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=HEART;		
RX	MEDLINE=21099252;	PubMed=11179219;	

RESULT 21	Q9VVT4	PRELIMINARY;	PRT;	336 AA.
ID	Q9VVT4			
AC	Q9VVT4			
DT	01-MAY-2000 (TrEMBLrel. 13, Created)			
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)			
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)			
DE	CG14074 protein.			
GN	CG14074			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephygrotidae; Drosophilidae; Drosophila.			
ON	NCBI_TaxID=7227;			
OX	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=BERKELEY;			
RX	MEDLINE=20196006; PubMed=10731132;			
RA	Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,			
RA	Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,			
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,			
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,			
RA	Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champagne M., Pfeiffer B.D.,			
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,			
RA	Abriel J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,			
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,			
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,			
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,			
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,			
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,			
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,			
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,			
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,			
RA	Fosler K., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,			
RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,			
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,			
RA	Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,			
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,			
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,			
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,			
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,			
RA	Merkulov G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,			
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,			
RA	Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,			
RA	Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,			
RA	Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,			
RA	Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,			
RA	Spirer E., Spradling A.C., Stapleton M., Strong R., Sun E.,			
RA	Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,			
RA	Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,			
RA	Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,			
RA	Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,			
RA	Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,			
RA	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;			
RT	"The genome sequence of Drosophila melanogaster."			
RL	Science 287:2185-2195(2000).			
DR	EMBL: AE003519; AAF49224.1;			
DR	FlyBase: FBgn0036818; CG14074.			
DR	InterPro: IPR002453; Beta_tubulin.			
DR	PROSITE: PS00228; TUBULIN_B.AUTOREG.1.			
SQ	SEQUENCE 336 AA; 38084 MW; 266856C3D947A30A CRC64;			
Query Match	9.7%; Score 97.5; DB 5; Length 336;			
Best Local Similarity	30.1%; Pred. No. 2;			
Matches	22; Conservative 17; Mismatches 29; Indels 1;			
QY	115 PRITKFDDEPHKRRKRS-----KSQSSKSQTENNRSSQSNAAHFTSKKRDTRKROER 169			
Db	79 PRQSVDRSSRRSSRRSHKAKKRRRTSSRRSSRRSSRRSSRRSSRRKRRKKKKVKK 138			
QY	170 HIKEEQDKWTS 182			
Db	139 HKSHRRRRSSQS 151			

	Best Local Similarity	34.3%;	Pred. No.	2.3;	Gaps	
	Matches	24;	Conservative	17;	Mismatches	21; Indels
Qy	122	DERPHERRKSRKSQS-----SKSQTRNNRSQSNANAHFTSKKRDKTK--RRQRHIKE	173			
		I:: I::				
Db	248	DKKEERHHSRKSRSRRRTTSSSRHRRSRSRRFSHKSRSRRRSKSPRRRRSHRE	307			
Qy	174	EQDKEHTSAK	183			
		: : : :				
Db	308	RGRRSNSTSK	317			
		:				
RESULT	18					
Q17595						
ID	Q17595	PRELIMINARY;	PRT;	535 AA.		
AC	Q17595;					
DT	01-NOV-1996;	(TrEMBLrel. 01, Created)				
DT	01-NOV-1996;	(TrEMBLrel. 01, Last sequence update)				
DT	01-DEC-2001;	(TrEMBLrel. 19, Last annotation update)				
DE	Hypothetical 60.3 kDa protein.					
GN	C02H7.1					
OS	Caenorhabditis elegans.					
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;					
OC	Rhabditidae; Pelodierinae; Caenorhabditis.					
OX	NCBI_TaxID=6239;					
ON	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=BRI-STOL N2;					
RX	MEDLINE=99369613; PubMed=9851916;					
RA	None;					
RT	"Genome sequence of the nematode C. elegans: a platform for					
RL	investigating biology. The C. elegans Sequencing Consortium."					
RL	Science 282:2012-2018(1998).					

```

RN      | [2]
RP      | SEQUENCE FROM N.A.
RC      | STRAIN-BRISTOL N2;
RA      | Leimbac D., Minx M.;
RT      | "the sequence of C. elegans cosmid C02H7.";
RL      | Submitted (FEB-1996) to the EMBL/GenBank/DDBJ databases.
RN      | [3]
RP      | SEQUENCE FROM N.A.
RC      | STRAIN=BRISTOL N2;
RA      | Waterston R.;
RT      | "Direct Submission.";
RL      | Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR      | EMBL; U99945; AAC47924.1; -.
KW      | Hypothetical protein.
SQ      | SEQUENCE   535 AA;  60303 MW;  03B2D8EBE43DFFB6 CRC64;

Query Match |          9.8%; Score 99; DB 5; Length 535;
Best Local Similarity 19.0%; Pred.No. 2.5;
Matches 40; Conservative 39; Mismatches 87; Indels 44; Gaps

Qy      | 1 MRKEVTEPLNLYNKYPGPQPIHFENIVKSDDIIEFQLVIN EKSAFDVTVFQGREFSEILKY 60
        | | : :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db      | 1 MSVEETREIL-----EKVIQKPQLTDQLLSRPPFPKFIVDI----VSNVIKST 43

Qy      | 61 DFIVGWCNGEQLRLRGFYKDASTIRNSRISLED----YIK----- 98
        | :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db      | 44 GYLKTDFTDDEIKSAGNDKNKTAFDLKLKI LDGSLKNVKAAKTIISCKDAEETNMQLQ 103

Qy      | 99 ----EYCFFGCAYFVLEPNPDKIDDERPHKRKRSKSQSOSKTSOTRRNNRSQSNANAH 154
        | | : :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db      | 104 MLGTNATSFNSRNGTGEKKKKVKKKEDKKGDDEBKSTTKRKSRKKKETHEEKESKKKS 163

Qy      | 155 FTSK-RKKDTKRRQRHHIKEEQDKEMTSK 183
        | | : :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db      | 164 AEEKERKKKSSSKERHKHSDDRSSESSEK 193

RESULT 19
Q13427 ID Q13427 PRELIMINARY; PRT; 754 AA.
```



```
DR EMBL; Z99281; CAB54458.1; .
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PS50102; RRM; 1.
SQ SEQUENCE 659 AA; 74740 MW; 8FC6694B17D98989 CRC64;

Query Match 10.4%; Score 105; DB 5; Length 659;
Best Local Similarity 23.4%; Pred. No. 1;
Matches 46; Conservative 24; Mismatches 93; Indels 34; Gaps 5;

Qy 11 NYNKYPGQFIHFENIVKSDDI-----EFQLVINE-----KSAFDV 46
    ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 22 NYNNNNHHHNNLPRINETQVPSKHTFIRGLHGDIISTEEIKVIGKVGKISDF 81
    ||| : : : : : : : : : : : : : : : : : : : : : : : :

Qy 47 TVFGQRFSEILLKYDFIVDGMGNEQLRLRGFYKDASTRKNSRISRLEDYIKCYNFCFA 106
    ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 82 VKVAQDKSKIFVAVFENRDEAKEFMEF---YSDREFMGCRCDLSWFRD-IRRYCAYORA 137
    ||| : : : : : : : : : : : : : : : : : : : : : : : :

Qy 107 YFVLENPNPROIKFDFRPHKR-----RKSRKSSQSSKQSTRNNRSOSNANAHFTSKR 160
    ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 138 KOVRSNSQRRRRSDSQSSKRSASPPVVRGSKSRSRSPSRSSRSRSRSVSRKSP 197
    ||| : : : : : : : : : : : : : : : : : : : : : : : :

Qy 161 KDKRRQRERHIKEQDK 177
    ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 198 ASRNGRVSRRSRSSQNR 214
    ||| : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 11
Q9D8S5 ID Q9D8S5 PRELIMINARY; PRT; 270 AA.
AC Q9D8S5;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Splicing factor, arginine/serine-rich 5 (SRp40, HRS).
GN SFRS5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=pancreas;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojancic N., Carninci P., de Bonaudo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
DR EMBL; AK007729; BAB25217.1; .
DR MGD; MGI:98287; Sfrs5.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 2.
DR SMART; SM00360; RRM; 2.
DR PROSITE; PS50102; RRM; 2.
DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
SQ SEQUENCE 270 AA; 30978 MW; EA3E8C9E5D10B680 CRC64;

Query Match 10.0%; Score 101; DB 4; Length 508;
Best Local Similarity 27.4%; Pred. No. 1.6;
Matches 43; Conservative 26; Mismatches 68; Indels 20; Gaps 6;

Qy 32 IEF--QLVINEKSAFDVTVFGQRFSEILLKYDFIVDGMGNEQLRLRGFYKDASTRKNSR 89
    ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 111 VEFALQNSVPRALAFNGVMGDFRPLKINHNSNAIVKP---PEMTPQAAKELEVWKR-- 165
    ||| : : : : : : : : : : : : : : : : : : : : : : : :

Qy 90 ISRLEDYIKCYNFCGAYFVLENPNPROIKFDFRPHKRKRKSRKSSQSSKQSTRNNRSQS 149
    ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 166 -----VREAQSFISAIEPEGSKNKRGRSRSHRSKRS---SKSHSRKRKRSQS 215
    ||| : : : : : : : : : : : : : : : : : : : : : : : :

Query Match 10.4%; Score 104.5; DB 11; Length 270;
Best Local Similarity 21.1%; Pred. No. 0.43;
Matches 44; Conservative 44; Mismatches 72; Indels 49; Gaps 8;

Qy 20 FIHFENIVKSDDIIEFQLVINEKSAFDVTV-----FGQRFSEILLKYD--- 61
    ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 41 FVEFEIPRDADAVVYELDGKELCSERVTEHARARSGRGRYSDRFSRRPRNDRN 100
    ||| : : : : : : : : : : : : : : : : : : : : : : : :

Qy 62 -----FIVGD-----WNEQLRLRGFYKDASTR-KNSRISRLEDYIKCYNFCG 105
    ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 101 APPVRTENRLIVENLSRVSWQD-----LKDFMRQAGEVTFADAHPRKLNQGVVEFASVD 156
    ||| : : : : : : : : : : : : : : : : : : : : : : : :

Qy 106 AVFVLJ-----NPNPRDIK-FDDERPHKRKRKSRKSSQSS-----SKSOTNNRSQSNANAH 154
    ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 157 LKNAIKLUSGKEINGRKIKLIEGSKRHSRSRSTRSSRSRKRKSRKSRKSR 216
    ||| : : : : : : : : : : : : : : : : : : : : : : : :

Qy 155 FTSKKIKDKTKRQRERHIKEEQDKEMTSK 183
    ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 217 SRSRSRKRKSRSGSRSPVPEKSKRGSSSR 245
    ||| : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 12
Q8WXA9 ID Q8WXA9 PRELIMINARY; PRT; 508 AA.
AC Q8WXA9;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Splicing factor, arginine/serine-rich 12.
GN SFRS12.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=pancreas;
RA Zhang D.L., Sun X.J., Ling L.J., Chen R.S., Ma D.L.;
RT "Molecular cloning, characterization, chromosomal assignment and
RT verification of SFRS12 (SRp308), a novel member of human SR protein
RT superfamily and a human homolog of rat SRp86 (Poster Abstracts 4,
RT poster #241).";
RL (In) Unknown A. (eds.);
RL HUGO Seventh International Human Genome Meeting, pp.244-244, Shanghai,
RL China (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=pancreas;
RA Zhang D.L., Sun X.J., Ling L.J., Chen R.S., Ma D.L.;
RT "Molecular cloning, characterization, chromosomal assignment, genomic
RT organization and verification of SFRS12 (SRp308), a novel member of
RL human SR protein superfamily and a human homolog of rat SRp86.";
DR EMBL; AF459094; AAL67778.1; .
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PS50102; RRM; 1.
SQ SEQUENCE 508 AA; 59380 MW; 878DC088B642EC3F CRC64;
```



```
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the gram-positive bacterium Bacillus
RL subtilis."
RN Nature 390:249-256(1997).
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z99120; CAB15221.1; -.
KW Complete proteome.
SQ SEQUENCE 102 AA; 12155 MW; 0551F0BF8AEF3275 CRC64;

Query Match 23.0%; Score 232.5; DB 16; Length 102;
Best Local Similarity 51.7%; Pred. No. 3.7e-12;
Matches 46; Conservative 16; Mismatches 26; Indels 1; Gaps 1;

QY 24 ENIVKSDIEFQVINEKSAFDTVFQRESEILLKYDFIVGDMGNEQLRLGRFYKRDAS- 82
DB 10 EIMILQNAEFELVHNPKDGFNEAFKARYSDILNKYDIVGDMGYGOLRLKGFDDQNG 69
QY 83 TIRKNSRISLEDYIKRYNFCGAYFVLE 111
DB 70 KATFEKISTLDEVYIYECNFCGAYFVLK 98

RESULT 8
Q8TIT2
ID Q8TIT2 PRELIMINARY; PRT; 1080 AA.
AC Q8TIT2;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 126 kDa protein.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT "Sequence and Analysis of Chromosome 2 of Dictyostelium."
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC116102; AAL93605.1; -.
KW Hypothetical protein.
SQ SEQUENCE 1080 AA; 126556 MW; 432C39D84C26ED29 CRC64;

Query Match 10.6%; Score 106.5; DB 5; Length 1080;
Best Local Similarity 18.1%; Pred. No. 1.3;
Matches 37; Conservative 51; Mismatches 71; Indels 45; Gaps 7;

QY 1 MRKEVTEPEMLNKNYKPGQPIHFENI-----VKSDIEFQV----- 38
DB 505 LRKEIQKLDIEKYAKP---EETERYKLLVQRDVESKNIEINKLLEKINITIKONND 560
QY 39 ---NEKSAFDTVFQRFSEILLKYDFIVGDMGNEQLRLGRFYKRDASTIRKNSRISLED 95
DB 561 LKNEQQL-----KQSELTLSNLKSNKN--ININNLKDN--NIEDNSKIKQQQE 610
QY 96 YIKRYCNFCGAYFVLEPNRPDIKFDDEPHKRRKRSKSKSQSQRNRSQSNAHAF 155
DB 611 QQQQ-----OHOKIDNNKKEEIKKQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 664
QY 156 TSKRKDKTKRQERHIKEEDKEM 179
DB 665 QQQQQQQQQQQQQQQEQEKDKEM 688

RESULT 9
Q9U203
ID Q9U203 PRELIMINARY; PRT; 553 AA.
```

```
AC Q9U203;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Y57G11C.9a protein.
GN Y57G11C.9A.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA McMurray A.A.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL; Z99281; CAB54457.1; -.
DR InterPro: IPR000504; RNA_rec_mot.
DR Pfam: PF00076; rrm; 1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PS50102; RRM; 1.
SQ SEQUENCE 553 AA; 63571 MW; A7F22D32D8E99C36 CRC64;

Query Match 10.4%; Score 105; DB 5; Length 553;
Best Local Similarity 23.4%; Pred. No. 0.84;
Matches 46; Conservative 24; Mismatches 93; Indels 34; Gaps 5;

QY 11 NNYKPGQPIHFENIVKSDI-----EFQLVINE---KSAFDV 46
DB 22 NNNNNHHHNNLPINETQVPQSKHTIFIRGLHGDISTEEIKYIGEKVKISFDF 81
QY 47 TVFGQRESEILLKYDFIVGDMGNEQLRLGRFYKRDASTIRKNSRISLEDYIKRYCNFCGA 106
DB 82 VKVAQDKSKIFVAVRFENDEAKEFMET---YSDREFMGRCGLSWPRD--IRRYCAYQRA 137
QY 107 YFVLENPNDIRKFDDEPHKRR-----RKSRKSKSKSQSQRNRSQSNAHAFTSKKR 160
DB 138 KQVRSNSQRRRRSDSQESSKRSASPPVRKGRSKRSRSPSRSRSRSRSRSRSRSKSP 197
QY 161 KDKRRKQERHIKEEDQK 177
DB 198 ASRNGRVSRHRSRQNR 214

RESULT 10
Q9U202
ID Q9U202 PRELIMINARY; PRT; 659 AA.
AC Q9U202;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Y57G11C.9b protein.
GN Y57G11C.9B.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA McMurray A.A.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology."
RL Science 282:2012-2018(1998).
```

```

Qy 26 IVKSDIEFQVINEKSAFDVTFVGORFSEILKYDFIVGDMGNEQLRGLRGFYKDS--TI 84
Db 1 MIKVDQHFELLENRECFNEQFARVSDILDKDYIVGDIYQDLRLKGFYKDSNKA 60

Qy 85 RNSRISRLDYKEYCNFGCAYFVLEPNPNRDKDFDDERPH 126
Db 61 EMSKRFNSIQDYIFEYCNFGCPFYVLRHLRSKQEVKRLTEEVH 102

RESULT 5
ID Q928N1 PRELIMINARY; PRT; 91 AA.
AC Q928N1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein lmo2402.
GN LMO2402 OR LIN2501.
OS Listeria monocytogenes, and
OC Listeria innocua.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1639, 1642;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=L.monocytogenes, and L.innocua;
RX STRAIN=EGD-E / SEROVAR 1/2A, AND CLIP 11262 / SEROVAR 6A;
MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
BA Quercio F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetoui F., Couve E., de Daruvar A., Dehoux P.,
RA Donann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkat G.,
RA Madoeno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordstiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL591983; CAD00480.1; -
DR EMBL; AL596172; CAC97728.1; -
DR ListList; LMO2402; -
DR ListList; LMO2402; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 91 AA; 11001 MW; AEE305DCCE28371C CRC64;

Query Match 23.7%; Score 239.5; DB 16; Length 91;
Best Local Similarity 51.2%; Pred. No. 8.6e-13;
Matches 43; Conservative 22; Mismatches 14; Indels 5; Gaps 2;

Qy 31 DIFQVINEKSAFDVTFVGORFSEILKYDFIVGDMGNEQLRGLRGFYKDS--TI 87
Db 7 DLNIEITNYDAFDEERLNFSDILGRDYIVGDMGYDQLRLKGFPEDDN--RKAAYD 64

Qy 88 SRISRLDYKEYCNFGCAYFVLE 111
Db 65 NKISLKEITYEYCNFGCAYFVIR 88

RESULT 6
ID Q9K7D1 PRELIMINARY; PRT; 93 AA.
AC Q9K7D1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein BH3433.
GN BH3433.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;

```

```

OC Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL; AP001518; BAB07152.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 93 AA; 11230 MW; 1AD87F55CA69E6E1 CRC64;

Query Match 23.2%; Score 234.5; DB 16; Length 93;
Best Local Similarity 49.4%; Pred. No. 2.3e-12;
Matches 44; Conservative 21; Mismatches 19; Indels 5; Gaps 2;

Qy 26 IVKSDIEFQVINEKSAFDVTFVGORFSEILKYDFIVGDMGNEQLRGLRGFYKDS--TI 82
Db 1 MVRISNMQFEVIENKDGWNEEAFVRYSDVLNKYDIYVDMGYNQLRLGFFDQNKKS 60

Qy 83 TIRKNSRISRLDYKEYCNFGCAYFVLE 111
Db 61 TYDK--KISTLDYLYEYCNFGCAYFVIR 87

RESULT 7
ID Q32127 PRELIMINARY; PRT; 102 AA.
AC Q32127;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE YUTD protein.
GN YUTD.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OX Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glasser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudaga B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serron P., Shin B.S., Soldo E.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpst P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler H., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,

```

```
QY 121 DDERPHKRRKS-RSKSQ-----SSKSQTRNNRSQSNANAHFTSKRRKDKTKR 165
Db 129 EEROPRRKKSPPKSKSRNRKPNYGNQOPATPKSKSK-RASKEKQEPENQAFTSKRRSNTK 187
QY 166 ROERHIKEQDKEMTSAKOHLFVRKN 192
Db 188 HKEKS-KRNQTSQNLTKSHFIIRKDK 213

RESULT 2
Q97RN6 PRELIMINARY; PRT; 170 AA.
AC Q97RN6;
DT 01-OCT-2001 (TReMBLrel. 18, Created)
DT 01-DEC-2001 (TReMBLrel. 18, Last sequence update)
DE Hypothetical protein SP0767.
GN SP0767.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1313;
RN 1
RP SEQUENCE FROM N.A.
RC STRAIN=TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tetelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Unayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapfel E., Khouri H., Wolf A.M., Utterback T.R., Hansen, C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae.";
RL Science 293:498-506(2001).
DR EMBL; AE007384; AAK74905.1; -.
DR TIGR; SP0767; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 170 AA; 20318 MW; 80273AB3AD61705E CRC64;

Query Match 43.3%; Score 436.5; DB 16; Length 170;
Best Local Similarity 46.1%; Pred. No. 7.9e-29;
Matches 88; Conservative 28; Mismatches 54; Indels 21; Gaps 2;

QY 1 MRKEVTPMLNKNYPGPGQFIHFENIVKSDDIEFQLVINEKSAFDVTVFGQRFSEILLKY 60
Db 1 MRKEIAPELYNKNYPGPGPEPHLHGDKVETEGIAFSLVENIKDAEDVTFNQRFFSEVLTKF 60

QY 61 DFIVGDMGNEQLRGLRGFYKDASTIRKSRISRLEDYIKEYCNFCAYFVLENPNRDIKF 120
Db 61 DYIVGDMNSNEQLRGLRGFYKDDTRTEKLEKISRLQDYLLEYCSYCAVFLNEAPKRAFS 120

QY 121 DDERPHKRRKSRSKSQSKSQSTRNNRSQSNANAHFTSKRRKDKTKRQERHTEKQDKEMT 180
Db 121 D-----KMRKTEETPSRK-----GKKPTQTRKRSNADKKNRRRQKDD 159

QY 181 SAKOHLFVRK 191
Db 160 KGQRHFVIRQK 170

RESULT 3
Q9CJ28 PRELIMINARY; PRT; 251 AA.
AC Q9CJ28;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-MAR-2001 (TReMBLrel. 17, Last sequence update)
DE Hypothetical protein ybiC.
GN ybiC OR LI0178.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
```

```
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN 1
RP SEQUENCE FROM N.A.
RC STRAIN=111403;
RX MEDLINE=212355186; PubMed=11337471;
RA Bolotin A., Wincker P., Mauger S., Jaillon O., Malarme K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403.";
RL Genome Res. 11:731-753(2001).
DR EMBL; AE006255; AAK04276.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 251 AA; 30095 MW; 975D661F0C97E01D CRC64;

Query Match 32.9%; Score 331.5; DB 16; Length 251;
Best Local Similarity 34.4%; Pred. No. 6e-20;
Matches 75; Conservative 38; Mismatches 64; Indels 41; Gaps 5;

QY 1 MRKEVTPMLNKNYPGPGQFIHFENIVKSDDIEFQLVINEKSAFDVTVFGQRFSEILLKY 60
Db 1 MAKVIDESKLNKNYPGGEHVMAGVEVVQVQGTQTHVHNVREAFDAEKLEQRFSDVLDKY 60

QY 61 DFIVGDMGNEQLRGLRGFYKDASTIRK-----NSRISRLEDYIKEYCNFCAYFVLENPNR 116
Db 61 DYIVGDMGFEQLRLKGFF---STSRRLADNKKIDHLEDYVNEVCNYGCAVFLRIRTK 117

QY 117 DIKFDDEP-----PHKRRKRSKSSQSSKTSOTNNRSQSNANAHFTSKRRKD 162
Db 118 DEAFVSEKLFTEKELKOGFKPRKRNRNRNRNWARDEOKVTKEDKRSEN-----SSEARKD 172

QY 163 TKRQERH-----IKEQDKREMTSAKQH 185
Db 173 FKIREKSTDRKPKVTDRNKNKVSQSKSQERTDNKKQN 210

RESULT 4
Q99VF1 PRELIMINARY; PRT; 127 AA.
AC Q99VF1;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Hypothetical protein SAV0925.
GN SAV0925 OR SA0786.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OS Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Staphylococcus.
OX NCBI_TaxID=158878; 158879;
RN 1
RP SEQUENCE FROM N.A.
RC SPECIES=S. aureus (strain Mu50), and S. aureus (strain N315);
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-Q., Ito T.,
RA Kanamori M., Matsunaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus.";
RL Lancet 357:1225-1240(2001).
DR EMBL; AP003360; BAB57087.1; -.
DR EMBL; AP003131; BAB42025.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 127 AA; 15401 MW; 4901EFB0676AE64E CRC64;

Query Match 23.9%; Score 241.5; DB 16; Length 127;
Best Local Similarity 46.1%; Pred. No. 8.4e-13;
Matches 47; Conservative 22; Mismatches 32; Indels 1; Gaps 1;
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2003, 09:51:02 ; Search time 31 seconds
(without alignments)
1276.163 Million cell updates/sec

Title: US-09-868-352-23
Perfect score: 1009
Sequence: 1 MRKEVTPMLNKNYPGPQF.....EEDDKEMTSKQHLFVRKN 192

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Length	ID	Description
1	558.5	55.4	214	16 Q99YU4	Q99YU4 streptococc
2	436.5	43.3	170	16 Q97RN6	Q97RN6 streptococc
3	331.5	32.9	251	16 Q9CJ28	Q9CJ28 lactococcu
4	241.5	23.9	127	16 Q99VF1	Q99VF1 staphylococ
5	239.5	23.7	91	16 Q92BN1	Q92BN1 listeria mo
6	234.5	23.2	93	16 Q9K7D1	Q9K7D1 bacillus ha
7	222.5	23.0	102	16 Q32127	Q32127 bacillus su
8	106.5	10.6	1080	5 Q8T1R2	Q8T1R2 dictyosteli
9	105	10.4	553	5 Q9U203	Q9U203 caenorhabdi
10	105	10.4	659	5 Q9U202	Q9U202 caenorhabdi
11	104.5	10.4	270	11 Q9D8S5	Q9D8S5 mus musculu
12	101	10.0	508	4 Q8WXA9	Q8WXA9 homo sapien
13	101	10.0	508	5 Q96I46	Q96I46 plasmodium
14	99	9.8	334	4 Q96I27	Q96I27 homo sapien
15	99	9.8	357	4 Q96DG9	Q96DG9 homo sapien
16	99	9.8	365	11 Q9CT17	Q9CT17 mus musculu

17	99	9.8	491	4 Q9HCZ3	Q9HCZ3 homo sapien
18	99	9.8	535	5 Q17595	Q17595 caenorhabdi
19	99	9.8	754	4 Q13427	Q13427 homo sapien
20	98.5	9.8	494	11 Q9JKL7	Q9JKL7 rattus norv
21	97.5	9.7	336	5 Q9VVT4	Q9VVT4 drosophila
22	97	9.6	752	11 Q5O35	Q5O35 rattus norv
23	96.5	9.6	513	10 Q9LW95	Q9LW95 nicotiana t
24	96.5	9.6	2281	11 Q99P84	Q99P84 rattus norv
25	95.5	9.5	718	4 Q9NVQ0	Q9NVQ0 homo sapien
26	95	9.4	712	5 Q22859	Q22859 caenorhabdi
27	94.5	9.4	1650	5 Q77328	Q77328 plasmodium
28	94	9.3	339	11 Q9DBP1	Q9DBP1 mus musculu
29	94	9.3	462	10 Q9LQ35	Q9LQ35 arabidopsis
30	93	9.2	339	11 Q921K3	Q921K3 mus musculu
31	93	9.2	453	2 Q9ADR3	Q9ADR3 staphylococ
32	93	9.2	866	5 Q8T8N7	Q8T8N7 caenorhabdi
33	93	9.2	924	5 Q9GYS7	Q9GYS7 caenorhabdi
34	93	9.2	925	5 Q9BIC2	Q9BIC2 caenorhabdi
35	93	9.2	1294	5 Q77361	Q77361 plasmodium
36	91.5	9.1	173	4 Q9Z8E5	Q9Z8E5 homo sapien
37	91.5	9.1	451	16 P71070	P71070 bacillus su
38	91.5	9.1	2081	10 Q9LH98	Q9LH98 arabidopsis
39	90.5	9.0	1056	16 Q8REF7	Q8REF7 fusobacteri
40	90.5	9.0	1103	4 Q9URV3	Q9URV3 homo sapien
41	90.5	9.0	1609	4 Q9P212	Q9P212 homo sapien
42	90.5	9.0	1994	4 Q9HC53	Q9HC53 homo sapien
43	90.5	9.0	2303	4 Q9HBX6	Q9HBX6 homo sapien
44	90	8.9	462	10 Q9ZUF4	Q9ZUF4 arabidopsis
45	90	8.9	470	5 Q22919	Q22919 caenorhabdi

ALIGNMENTS

RESULT 1

Q99YU4	ID	Q99YU4	PRELIMINARY;	PRT;	214 AA.
AC	Q99YU4:				
DT	01-JUN-2001	(TREMBLrel. 17, Created)			
DT	01-JUN-2001	(TREMBLrel. 17, Last sequence update)			
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)			
DE	Hypothetical. protein Spy1534.				
GN	SPY1534.				
OS	Streptococcus pyogenes				
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;				
OC	Streptococcaceae; Streptococcus.				
OX	NCBI_TaxID=314;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=SF370 / ATCC 700294 / SEROTYPE M1;				
RX	MEDLINE=2112684; PubMed=11296296;				
RA	Ferretti J.F., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,				
RA	Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,				
RA	Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,				
RA	Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;				
RT	"Complete genome sequence of an M1 strain of Streptococcus pyogenes.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).				
KW	EMBL; AE006386; AAK34328.1; -				
DR	Hypothetical. protein; Complete proteome.				
SQ	SEQUENCE 214 AA; 25536 MW; B054E21499807357 CRC64;				

Query Match 55.4%; Score 558.5; DB 16; Length 214;
Best Local Similarity 51.7%; Pred. No. 7.8e-39;
Matches 107; Conservative 42; Mismatches 41; Indels 17; Gaps 4;

Qy	1	MRKEVTPMLNKNYPGPQFTHFENIVKNSAFDVTVFGRFSEILLKY 60	
Db	9	MKKFISVEMYNKFPQKFIHFEEYKAEIGDILLLEDVKNKNAEDTTSFGORYTEVLKY 68	
Qy	61	DFIVGNGNSQLRGYKDKASTIRKNSRSLRLEDYIKYCNFCGAYFVLENPNRDIKF 120	
Db	69	DYIVGNGNEQLRGYKDKASTIRKNSRSLRLEDYIKYCNFCGAYFVLENLPQDIKF 128	